

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 01:09:48 ; Search time 647 Seconds
(without alignments)
9967.477 Million cell updates/sec

Title: US-09-765-271-55

Perfect score: 2389

Sequence: 1 TTCTTACGAGTTGGAGCTGT.....TAAGTACGAAAAATAAAC 2389

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2552756 seqs, 1349719017 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 200 summaries

Database :	N.Geneseq_19Jun03:*
1:	/SIDSI/gcgdata/geneseq/geneqn-emb1/NA1980.DAT:*
2:	/SIDSI/gcgdata/geneseq/geneqn-emb1/NA1981.DAT:*
3:	/SIDSI/gcgdata/geneseq/geneqn-emb1/NA1982.DAT:*
4:	/SIDSI/gcgdata/geneseq/geneqn-emb1/NA1983.DAT:*
5:	/SIDSI/gcgdata/geneseq/geneqn-emb1/NA1984.DAT:*
6:	/SIDSI/gcgdata/geneseq/geneqn-emb1/NA1985.DAT:*
7:	/SIDSI/gcgdata/geneseq/geneqn-emb1/NA1986.DAT:*
8:	/SIDSI/gcgdata/geneseq/geneqn-emb1/NA1987.DAT:*
9:	/SIDSI/gcgdata/geneseq/geneqn-emb1/NA1988.DAT:*
10:	/SIDSI/gcgdata/geneseq/geneqn-emb1/NA1989.DAT:*
11:	/SIDSI/gcgdata/geneseq/geneqn-emb1/NA1990.DAT:*
12:	/SIDSI/gcgdata/geneseq/geneqn-emb1/NA1991.DAT:*
13:	/SIDSI/gcgdata/geneseq/geneqn-emb1/NA1992.DAT:*
14:	/SIDSI/gcgdata/geneseq/geneqn-emb1/NA1993.DAT:*
15:	/SIDSI/gcgdata/geneseq/geneqn-emb1/NA1994.DAT:*
16:	/SIDSI/gcgdata/geneseq/geneqn-emb1/NA1995.DAT:*
17:	/SIDSI/gcgdata/geneseq/geneqn-emb1/NA1996.DAT:*
18:	/SIDSI/gcgdata/geneseq/geneqn-emb1/NA1997.DAT:*
19:	/SIDSI/gcgdata/geneseq/geneqn-emb1/NA1998.DAT:*
20:	/SIDSI/gcgdata/geneseq/geneqn-emb1/NA1999.DAT:*
21:	/SIDSI/gcgdata/geneseq/geneqn-emb1/NA2000.DAT:*
22:	/SIDSI/gcgdata/geneseq/geneqn-emb1/NA2001A.DAT:*
23:	/SIDSI/gcgdata/geneseq/geneqn-emb1/NA2001B.DAT:*
24:	/SIDSI/gcgdata/geneseq/geneqn-emb1/NA2002.DAT:*
25:	/SIDSI/gcgdata/geneseq/geneqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2389	100.0	2389	19	AAV27351
2	2389	100.0	2389	24	ABQ84819
3	2389	100.0	2451	21	AA047604
4	2338	97.9	8195	15	ABX06886
5	2338	97.9	8195	19	AAV52227
6	2338	97.9	2162598	25	ABSS56454
7	148	6.2	2523	21	AAA65731
8	148	6.2	2647	21	AAA65736

9	148	6.2	2647	24	ABK15103	DNA encoding Strep
10	53	2.2	504	21	AAA08556	S. pneumoniae 20 k
11	53	2.2	2163	20	AAV25394	Streptococcus pneu
12	53	2.2	2290	19	AAV27356	Streptococcus pneu
13	53	2.2	2359	24	ABQ84824	S. pneumoniae SP04
14	53	2.2	2457	25	AAV52376	Streptococcus pneu
15	53	2.2	2457	25	ABX06885	S. pneumoniae type
16	53	2.2	2481	21	AAA08557	S. pneumoniae 92 k
17	53	2.2	2481	21	AAA05417	Streptococcus pneu
18	53	2.2	2517	25	ABX06705	S. pneumoniae type
19	53	2.2	2531	21	AAA47602	Recombinant Varian
20	53	2.2	2531	21	AAA47605	Recombinant Varian
21	53	2.2	2639	21	AAA65737	Streptococcus pneu
22	53	2.2	2639	21	ABK15104	DNA encoding Strep
23	53	2.2	2162598	25	ABSS56454	Streptococcus pneu
24	50	2.1	492	20	AAV25393	Streptococcus pneu
25	44	1.8	973	19	AAV52488	Streptococcus pneu
26	38	1.6	3171	21	AAA65739	Streptococcus pneu
27	30	1.3	40	19	AAV27492	Streptococcus pneu
28	30	1.3	40	21	AAA47601	Primer for amplify
29	30	1.3	40	24	ABQ84960	Streptococcus pneu
30	27	1.1	36	21	AAA47599	Primer for amplify
31	27	1.1	37	19	AAV27491	Streptococcus pneu
32	27	1.1	37	24	ABQ84959	Streptococcus pneu
33	26	1.1	35	21	AAA47600	Primer for amplify
34	25	1.0	33	21	AAA65762	Streptococcus pneu
35	25	1.0	33	24	ABK33684	S. pneumoniae BVH-
36	25	1.0	34	21	AAA65764	Streptococcus pneu
37	25	1.0	34	24	ABK33686	S. pneumoniae BVH-
38	25	1.0	35	24	ABK33719	S. pneumoniae BVH-
39	24	1.0	1342	19	AAV27414	Streptococcus pneu
40	24	1.0	1442	24	ABQ84882	Streptococcus pneu
41	24	1.0	1398	24	ABN656839	Streptococcus poly
42	24	1.0	1455	21	AAA65733	Streptococcus pneu
43	24	1.0	1455	21	AAA47603	Recombinant Varian
44	24	1.0	1455	21	AAZ05473	Streptococcus pneu
45	24	1.0	1455	21	AAZ05484	Streptococcus pneu
46	24	1.0	2528	21	AAA65738	Streptococcus pneu
47	24	1.0	2528	24	ABK15105	DNA encoding Strep
48	24	1.0	3117	25	ABX06706	S. pneumoniae type
49	24	1.0	3120	21	AAA65730	Streptococcus pneu
50	24	1.0	5048	21	AAA65735	Streptococcus pneu
51	24	1.0	5048	24	ABK15101	DNA encoding Strep
52	24	1.0	6867	19	AAV52325	Streptococcus pneu
53	24	1.0	2155561	24	ABN71527	Streptococcus poly
54	22	0.9	1547	21	AAV37581	Streptococcus poly
55	21	0.9	29	21	AAA65761	Arbidolopsis thalia
56	21	0.9	29	24	ABK33683	S. pneumoniae BVH-
57	21	0.9	462586	25	ABQ84281	S. pneumoniae 2 Asthm
58	20	0.8	501	24	AAI69344	C. albicans hyphal
59	20	0.8	1146	21	AAA05814	Group B Streptococ
60	20	0.8	2466	24	ABN65935	Streptococcus poly
61	20	0.8	2466	24	ABN70334	Streptococcus poly
62	20	0.8	2469	21	AAA65740	Streptococcus pneu
63	20	0.8	2469	21	AAA05811	Group B Streptococ
64	20	0.8	2469	22	AAV54820	Streptococcus agal
65	20	0.8	2472	21	AAA65741	Streptococcus pneu
66	20	0.8	2475	24	ABN66838	Streptococcus poly
67	20	0.8	2478	22	AAV50036	Streptococcus pyog
68	20	0.8	5158	24	AAI69348	C. albicans contig
69	20	0.8	5215	20	AAV91105	Group B Streptococ
70	19	0.8	164	25	ABX54820	Bovine Est associa
71	19	0.8	426	23	ABV02215	Human prostate exp
72	19	0.8	427	22	AAV30789	Human cDNA encodin
73	19	0.8	468	23	ABV11384	Human prostate exp
74	19	0.8	481	23	ABV13259	Human prostate exp
75	19	0.8	481	23	ABV41454	Human prostate exp
76	19	0.8	693	24	ABV089355	Human prostate exp
77	19	0.8	1203	23	ABV21401	Human prostate exp
78	19	0.8	1205	23	AAV27219	Human prostate exp
79	19	0.8	1455	21	AAV05838	Human B Streptococ
80	19	0.8	1650	21	AAA70231	Plasmodium falciipa
81	19	0.8	7492	22	AAV90510	Mouse factor VIII

[illegible]

EN M09818930-A2.
XX 07-MAY-1998.
XX 30-OCT-1997; 97MO-US19422.
XX 31-OCT-1996; 96US-0029960.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Chou GH, Hromockyj A, Johnson LS, Kunsch CA;
XX WPI; 1998-272224/24.
XX P-PSDB; AAMS5090.
PT Nucleic acid encoding antigenic peptide(s) from Streptococcus
PT pneumoniae - or their epitope-containing fragments, useful in
PT protective or therapeutic vaccines, and for diagnosis
PS
PS Claim 1; Page 59; 118pp; English.
CC The present sequence encodes a protein from Streptococcus pneumoniae.
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein
CC can be useful in vaccines for inducing protective antibodies against
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
CC are used to detect Streptococcus infection (by usual hybridisation or
CC amplification methods), also for isolating Streptococcus genes or their
CC allelic variants. The protein can be used similarly to detect specific
CC antibodies in standard immunoassays, especially for diagnosing or
CC monitoring infections. Antibodies which bind the protein are used to
CC detect corresponding antigens, to purify the protein and for passive
CC immunisation (optionally coupled to a toxin). Vaccines are administered,
CC e.g. by injection, orally or through the skin, typically at 0.01-1000
CC (especially 10-300) mu g/ml per dose.
XX
XX Sequence 2389 BP; 830 A; 461 C; 486 G; 611 T; 1 other;
SQ
Query Match 100.0%; Score 2389; DB 19; Length 2389;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 481 AGAAGCTATATACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 540
DB 481 AGAAGCTATATACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 540
QY 541 TGTGATGCTTATATCGTTCTCATGAGATCATTTACATTAATTCCTAAGAAATGAGTT 600
DB 541 TGTGATGCTTATATCGTTCTCATGAGATCATTTACATTAATTCCTAAGAAATGAGTT 600
QY 601 ATCAGCTAGCGAGTTGGCTGTCGAGAAAGCCTTCTATCTGTGTCAGAAATCTGTCAA 660
DB 601 ATCAGCTAGCGAGTTGGCTGTCGAGAAAGCCTTCTATCTGTGTCAGAAATCTGTCAA 660
QY 661 TTCAAGAACTATGCGCGACAAAATAGGATTAACCTTCAAGAACTGGGTACTCTTC 720
DB 661 TTCAAGAACTATGCGCGACAAAATAGGATTAACCTTCAAGAACTGGGTACTCTTC 720
QY 721 TGTAAAGCAATCCAGAACTAACAATATCTAACAAGCAACAACACTAACAGTCA 780
DB 721 TGTAAAGCAATCCAGAACTAACAATATCTAACAAGCAACAACACTAACAGTCA 780
QY 781 AGCAAGTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
DB 781 AGCAAGTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
QY 841 TCAAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
DB 841 TCAAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
QY 901 AGCTAGAGTGTGAGTGCACACAGAGATCATTTACCTTACTCTTACTCTCAAT 960
DB 901 AGCTAGAGTGTGAGTGCACACAGAGATCATTTACCTTACTCTTACTCTCAAT 960
QY 961 GTCTGAATTTGGAAGCAAGATGCTGCTGATTAATCCCTGCTTCAACCAATG 1020
DB 961 GTCTGAATTTGGAAGCAAGATGCTGCTGATTAATCCCTGCTTCAACCAATG 1020
QY 1021 GGTACCAATTTCAAGGCGAGAAACCAAGTCCAAAGCTCCGGAACCTGACGAG 1080
DB 1021 GGTACCAATTTCAAGGCGAGAAACCAAGTCCAAAGCTCCGGAACCTGACGAG 1080
QY 1081 CCCGCAACCTGACCAATCTTAAATATGATGATGATGATGATGATGATGATGATGAT 1140
DB 1081 CCCGCAACCTGACCAATCTTAAATATGATGATGATGATGATGATGATGATGATGAT 1140
QY 1141 ACGAAAGTTGGGAGAGATATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
DB 1141 ACGAAAGTTGGGAGAGATATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
QY 1201 GAAAGATTTACATCTGAAATCTGTTAAATCTTGAAGCAAGTATCAAAAACAAGAG 1260
DB 1201 GAAAGATTTACATCTGAAATCTGTTAAATCTTGAAGCAAGTATCAAAAACAAGAG 1260
QY 1261 TGTTCACACACTTAACTGCTAAAAAATGTTCTCTCTGTCGACCAAGATTTTA 1320
DB 1261 TGTTCACACACTTAACTGCTAAAAAATGTTCTCTCTGTCGACCAAGATTTTA 1320
QY 1321 TGAATTAACATTAATCTGTTAATGATGATGATGATGATGATGATGATGATGATGAT 1380
DB 1321 TGAATTAACATTAATCTGTTAATGATGATGATGATGATGATGATGATGATGATGAT 1380
QY 1381 TAAATTCGATTTCCAGGCTTGAACAATTTTGAAGCAAGTATGATGATGATGATGAT 1440
DB 1381 TAAATTCGATTTCCAGGCTTGAACAATTTTGAAGCAAGTATGATGATGATGATGAT 1440
QY 1441 TAAAGAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
DB 1441 TAAAGAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
QY 1501 ACTTGCAACCAATTTCTCAATTTGATGATGATGATGATGATGATGATGATGATGAT 1560
DB 1501 ACTTGCAACCAATTTCTCAATTTGATGATGATGATGATGATGATGATGATGATGAT 1560

```

QY 1561 AGCTGATAGATATCAACGTCAGATGTTTACATTTTGTGTAACATGATATATCAGTGA 1620
D 1561 AGCTGATAGATATCAACGTCAGATGTTTACATTTTGTGTAACATGATATATCAGTGA 1620
QY 1621 TGAAGGAGATGATATGTAAGGCTCATATGGCCATATGTCATCTGGATTTGAAAAAGATAG 1680
D 1621 TGAAGGAGATGATATGTAAGGCTCATATGGCCATATGTCATCTGGATTTGAAAAAGATAG 1680
QY 1681 CCTTTCGATAGAGAAAAAGTTGACGCTCAAGCCATATCTAAAGAAAAAGGATCTCTAC 1740
D 1681 CCTTTCGATAGAGAAAAAGTTGACGCTCAAGCCATATCTAAAGAAAAAGGATCTCTAC 1740
QY 1741 TCCATCTCCAGACGAGATGTTAAAGCAATCCAATGAGATGTGACGACGATATTTA 1800
D 1741 TCCATCTCCAGACGAGATGTTAAAGCAATCCAATGAGATGTGACGACGATATTTA 1800
QY 1801 CAATCTGTGAAAGGGGAAAAAGCAATTCCTGCTGTTGACTTCATATATGTTGAGCA 1860
D 1801 CAATCTGTGAAAGGGGAAAAAGCAATTCCTGCTGTTGACTTCATATATGTTGAGCA 1860
QY 1861 TACAGTTGAGGTTAAAAAGGTAATTTGATATTCCTCATAGGATTCATTCATATAT 1920
D 1861 TACAGTTGAGGTTAAAAAGGTAATTTGATATTCCTCATAGGATTCATTCATATAT 1920
QY 1921 TAAATTTGCTGTTGTTGATGATCAACATACAAAGCTCAATGCTATACCTTGAGAGA 1980
D 1921 TAAATTTGCTGTTGTTGATGATCAACATACAAAGCTCAATGCTATACCTTGAGAGA 1980
QY 1981 TTTGTTTGGACGATTAAGTACGTAAGAACACCTTGACGAGACGTCACATTCATATGA 2040
D 1981 TTTGTTTGGACGATTAAGTACGTAAGAACACCTTGACGAGACGTCACATTCATATGA 2040
QY 2041 TGGATGGGGCAATGCGACGATGAGTGTGTTGGGCAAGAACCCACAGTGAAGATCCAAA 2100
D 2041 TGGATGGGGCAATGCGACGATGAGTGTGTTGGGCAAGAACCCACAGTGAAGATCCAAA 2100
QY 2101 TAAAGACTTCAAGCGGATGAAAGAGCCAGTAGAGAGAAACACCTGTGAGCCAGAGTCCC 2160
D 2101 TAAAGACTTCAAGCGGATGAAAGAGCCAGTAGAGAGAAACACCTGTGAGCCAGAGTCCC 2160
QY 2161 TCAAGTAGAGACTGAAAAAGTGAAGGCCCACTCAAGAGCAGAGAGTTTGTTCGCA 2220
D 2161 TCAAGTAGAGACTGAAAAAGTGAAGGCCCACTCAAGAGCAGAGAGTTTGTTCGCA 2220
QY 2221 AGTAAAGGATCTGATGTAAGGCAATGCAAGCAAGAACTCTAGCTGTTTACGAAATTA 2280
D 2221 AGTAAAGGATCTGATGTAAGGCAATGCAAGCAAGAACTCTAGCTGTTTACGAAATTA 2280
QY 2281 TTTGACTCTTCAAAATTATGATTAACAATAGTATCATGSCAGAGAGAAAAATTACTTGC 2340
D 2281 TTTGACTCTTCAAAATTATGATTAACAATAGTATCATGSCAGAGAGAAAAATTACTTGC 2340
QY 2341 GTTGTAAAGGAAGTAATCTCTCATCTGTAAGTGAAGAAAAAATAAAC 2389
D 2341 GTTGTAAAGGAAGTAATCTCTCATCTGTAAGTGAAGAAAAAATAAAC 2389

```

RESULT 2

```

AB084819
ID AB084819 standard; DNA; 2389 BP.
AC
XX AB084819;
XX
DT 04-SEP-2002 (first entry)
XX
DE S. pneumoniae SP036 nucleotide sequence SEQ ID NO:55.
XX
KM Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
XX antibiotic; Streptococcal infection; detection; gene; ds.
OS Streptococcus pneumoniae.
XX
XX US2002061545-A1.
PN

```

```

XX 23-MAY-2002.
PD
XX 22-JAN-2001; 2001US-0765272.
PE
XX 30-OCT-1997; 97US-0961083.
PR
XX (CHOI/) CHOI G H.
PA (KUN/) KUNSC C A.
PA (BARA/) BARASH S C.
PA (DILL/) DILLON P J.
PA (DOUG/) DOUGHERTY B.
PA (FANN/) FANNON M R.
PA (ROSE/) ROSEN C A.
PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
PI Rosen CA;
DR WPI; 2002-479261/51.
DR P-PSDB; ABP54584.
XX
XX New Streptococcus pneumoniae antigens, useful for detecting
FT Streptococcus and for preventing or attenuating disease caused by
PT Streptococcus infection -
PS
XX Claim 1; Page 27; 70pp; English.

```

AB084792 to AB084904 represents nucleic acids which encode the Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S. pneumoniae antigens have antibacterial activity and can be used in vaccines. The S. pneumoniae antigens can also be used to prevent or attenuate a Streptococcal infection in an animal. The CC polynucleotides encoding the S. pneumoniae antigens can be used to detect Streptococcus nucleic acids. AB084905 to AB085130 represent primers used in the cloning of S. pneumoniae ORFs (open reading frames) which are used in an example from the present invention.

Sequence 2389 BP; 830 A; 461 C; 486 G; 611 T; 1 other;

Query Match 100.0%; Score 2389; DB 24; Length 2389;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 TTTCTTACGATGTTGGGACTGATCAAGCTAGAAACGTTAAGAAAAATATCGTGTCTTA 60
D 1 TTTCTTACGATGTTGGGACTGATCAAGCTAGAAACGTTAAGAAAAATATCGTGTCTTA 60
QY 61 TATAGATGAAAAACAAACGACGCAAAAAACGAGAAATTTGACTCTGATGAGTTAGCAA 120
D 61 TATAGATGAAAAACAAACGACGCAAAAAACGAGAAATTTGACTCTGATGAGTTAGCAA 120
QY 121 GCGTGAAGGAATCAATCTGAGCAAAATCGTCATCAATCAATACAGACCAAGGCTTGTCC 180
D 121 GCGTGAAGGAATCAATCTGAGCAAAATCGTCATCAATCAATACAGACCAAGGCTTGTCC 180
QY 181 TTTGATGAGGACACCATATCATTTTCAATGTTGTTTACATGTTTACATGTTTACATG 240
D 181 TTTGATGAGGACACCATATCATTTTCAATGTTGTTTACATGTTTACATGTTTACATG 240
QY 241 TGAAGATTTACTCATGAAAGATCCAAACTTAAGCTTAAAGATGAGATATGTTAATGA 300
D 241 TGAAGATTTACTCATGAAAGATCCAAACTTAAGCTTAAAGATGAGATATGTTAATGA 300
QY 301 GGTCAAGGTTGATATGTTTCAAGGATGATGAAATATCTATGTTTACCTTAAGATGC 360
D 301 GGTCAAGGTTGATATGTTTCAAGGATGATGAAATATCTATGTTTACCTTAAGATGC 360
QY 361 TGCCACGCGGATTAAGCTCGTACAAAGAGAAATCAATGACAAAGAAACAGAGCATAG 420
D 361 TGCCACGCGGATTAAGCTCGTACAAAGAGAAATCAATGACAAAGAAACAGAGCATAG 420
QY 421 TCAACATCGTGAAGGTGAATCCAGAAAGATGATGCTGTGCTTGGCAGCTTGGCA 480
D 421 TCAACATCGTGAAGGTGAATCCAGAAAGATGATGCTGTGCTTGGCAGCTTGGCA 480

```


Db 421 TCACATCGTGAAGGTGAACCTCAAGAAACGATGCTGTGCTTCCTTGCGACGTTTCGCA 480
 Qy 481 AGGAGCTTACTACAGATGATGTTATCTTTTATGCTTGATGATGAGGATAC 540
 Db 481 AGGAGCTTACTACAGATGATGTTATCTTTTATGCTTGATGATGAGGATAC 540
 Qy 541 TGGTATGCTTATATGCTTCTCATGAGATCATTTACATTCATCTTAAGAAATGAT 600
 Db 541 TGGTATGCTTATATGCTTCTCATGAGATCATTTACATTCATCTTAAGAAATGAT 600
 Qy 601 ATCAGTACGAGTGGCTGCTGCGAGAGCCTTCCTATGCTGAGAAATCTGTCAA 660
 Db 601 ATCAGTACGAGTGGCTGCTGCGAGAGCCTTCCTATGCTGAGAAATCTGTCAA 660
 Qy 661 TTCAGAACCTATCGCGCAAAATAGCATTAACCTTAAGAAACAACTGGGTACTTC 720
 Db 661 TTCAGAACCTATCGCGCAAAATAGCATTAACCTTAAGAAACAACTGGGTACTTC 720
 Qy 721 TGTAAAGCAATCCAGAACTACAAATCTAACAACAAACAAACAGCAACTAACAGTCA 780
 Db 721 TGTAAAGCAATCCAGAACTACAAATCTAACAACAAACAAACAGCAACTAACAGTCA 780
 Qy 781 AGCAAGTCAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
 Db 781 AGCAAGTCAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
 Qy 841 TCAAGCAATGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
 Db 841 TCAAGCAATGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
 Qy 901 AGCTAGAGTGTGCGATGCGACCAAGAGATCATTCACATTCATCTTCATCTCAAT 960
 Db 901 AGCTAGAGTGTGCGATGCGACCAAGAGATCATTCACATTCATCTTCATCTCAAT 960
 Qy 961 GTCTGAATGGAAGAAAGATGCTGCTGATGATGATGATGATGATGATGATGATGAT 1020
 Db 961 GTCTGAATGGAAGAAAGATGCTGCTGATGATGATGATGATGATGATGATGATGAT 1020
 Qy 1021 GGTACCAAGTTCAGAGGCGAAGCAACAGTCCGACCGAGCTGAGTCCAG 1080
 Db 1021 GGTACCAAGTTCAGAGGCGAAGCAACAGTCCGACCGAGCTGAGTCCAG 1080
 Qy 1081 CCCGCAACCTGACCAAAATCTTAAATGATCAATCTTCTGTTAGTCACTGAT 1140
 Db 1081 CCCGCAACCTGACCAAAATCTTAAATGATCAATCTTCTGTTAGTCACTGAT 1140
 Qy 1141 ACGAAAGTGGGGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
 Db 1141 ACGAAAGTGGGGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
 Qy 1201 GAAAGATTTACATCTGAAACCTGTTAAATCTTGAAGCAAGTTATCAAAACAAAGAG 1260
 Db 1201 GAAAGATTTACATCTGAAACCTGTTAAATCTTGAAGCAAGTTATCAAAACAAAGAG 1260
 Qy 1261 TGTTCACACACTTTACTGCTGTAAGAAAGAAATGTTGCTCTGCGACCAAGAAATTTTA 1320
 Db 1261 TGTTCACACACTTTACTGCTGTAAGAAAGAAATGTTGCTCTGCGACCAAGAAATTTTA 1320
 Qy 1321 TGATAAGCATTAATCTGTTAACTGAGGCTCATAAAGCTTGTGTTGAAATTAAGGCTG 1380
 Db 1321 TGATAAGCATTAATCTGTTAACTGAGGCTCATAAAGCTTGTGTTGAAATTAAGGCTG 1380
 Qy 1381 TAAATCTGATTTCCAAAGCTTGAACAAATTAATTAAGAAATTAAGAAATTAAGAAAT 1440
 Db 1381 TAAATCTGATTTCCAAAGCTTGAACAAATTAATTAAGAAATTAAGAAATTAAGAAAT 1440
 Qy 1441 TAAAGAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
 Db 1441 TAAAGAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
 Qy 1501 ACTTGCAACCAAAATTTCTCAATGATGATGATGATGATGATGATGATGATGATGAT 1560
 Db 1501 ACTTGCAACCAAAATTTCTCAATGATGATGATGATGATGATGATGATGATGATGAT 1560

Qy 1561 AGCTGATAGTATACAAAGTGCAGATGATGATGATGATGATGATGATGATGATGATGAT 1620
 Db 1561 AGCTGATAGTATACAAAGTGCAGATGATGATGATGATGATGATGATGATGATGATGAT 1620
 Qy 1621 TGAAGAGATGATATGTAACGCTCATATGAGGCAATGATGATGATGATGATGATGATGAT 1680
 Db 1621 TGAAGAGATGATATGTAACGCTCATATGAGGCAATGATGATGATGATGATGATGATGAT 1680
 Qy 1681 CCTTTCTGATTAAGAAAGTTCAGCTCAAGCTTATCTAAGAAAGAAAGTATCTTAC 1740
 Db 1681 CCTTTCTGATTAAGAAAGTTCAGCTCAAGCTTATCTAAGAAAGAAAGTATCTTAC 1740
 Qy 1741 TCCATCTCAGACGAGATGTTAAAGCAATCCAACTGAGATGATGATGATGATGATGAT 1800
 Db 1741 TCCATCTCAGACGAGATGTTAAAGCAATCCAACTGAGATGATGATGATGATGATGAT 1800
 Qy 1801 CAATGCTGTAAGAGGGAAGAAACGAATTCACCTGCTGACTTCCATATATGTTGAGCA 1860
 Db 1801 CAATGCTGTAAGAGGGAAGAAACGAATTCACCTGCTGACTTCCATATATGTTGAGCA 1860
 Qy 1861 TACAGTGAAGTTAAAGCGTAAATTTGATGATGATGATGATGATGATGATGATGAT 1920
 Db 1861 TACAGTGAAGTTAAAGCGTAAATTTGATGATGATGATGATGATGATGATGATGAT 1920
 Qy 1921 TAAATTTGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980
 Db 1921 TAAATTTGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980
 Qy 1981 TTTGTTGCGAGATTAAGTACTAGTAAACACCTGCGAGCAAGCTCCACATTTCAATGA 2040
 Db 1981 TTTGTTGCGAGATTAAGTACTAGTAAACACCTGCGAGCAAGCTCCACATTTCAATGA 2040
 Qy 2041 TGGATGGGCAATGCGAGTACATGTTGATGATGATGATGATGATGATGATGATGATGAT 2100
 Db 2041 TGGATGGGCAATGCGAGTACATGTTGATGATGATGATGATGATGATGATGATGATGAT 2100
 Qy 2101 TAAAGACTTCAAGGGGATGAAGAGCAGTAGAGAAACACCTGCTGAGCCGAGAGTCCC 2160
 Db 2101 TAAAGACTTCAAGGGGATGAAGAGCAGTAGAGAAACACCTGCTGAGCCGAGAGTCCC 2160
 Qy 2161 TCAAGTAGAGTCAAGAAAGTGAAGCCCACTCAAGAGAGAGAGAGTTCCTGCGAA 2220
 Db 2161 TCAAGTAGAGTCAAGAAAGTGAAGCCCACTCAAGAGAGAGAGAGTTCCTGCGAA 2220
 Qy 2221 AGTAACGATTTCTAGTCTGAAGCCCAATGCAACAGAACTCTAGCTGTTACGAATTA 2280
 Db 2221 AGTAACGATTTCTAGTCTGAAGCCCAATGCAACAGAACTCTAGCTGTTACGAATTA 2280
 Qy 2281 TTTGACTCTTCAAAATTATGATGAACAAATGATGATGATGATGATGATGATGATGATGAT 2340
 Db 2281 TTTGACTCTTCAAAATTATGATGAACAAATGATGATGATGATGATGATGATGATGATGAT 2340
 Qy 2341 GTTGTAAAGAGAAATCTTCACTGTAAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAG 2399
 Db 2341 GTTGTAAAGAGAAATCTTCACTGTAAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAG 2399

RESULT 3
 AAA47604
 ID AAA47604 standard; DNA; 2451 bp.
 XX
 AC AAA47604;
 XX
 DT 20-OCT-2000 (first entry)
 XX
 DE Recombinant variant of Sp36 gene (Sp36A) of *S. pneumoniae*.
 XX
 KW Streptococcus pneumoniae; infection; vaccine; coiled coil region;
 KW histidine triad residue; Sp36; antibody; otitis media;
 KW nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
 KW meningitis; lobar pneumonia; ds.
 XX

OS Streptococcus pneumoniae.
 XX Key Location/Qualifiers
 FT CDS 1.2451
 FT /tag= a
 FT /product= Sp36a polypeptide
 FN WO200037105-A2.
 XX 29-JUN-2000.
 XX 21-DEC-1999; 99WO-US30390.
 XX 21-DEC-1998; 98US-0113048.
 XX (MEDI-) MEDIMUNE INC.
 PA Johnson LS, Koenig S, Adamou JE;
 P1 WPI; 2000-452129/39.
 DR P-PSDB; AAB01468.
 XX
 PT Vaccine useful for prophylaxis and treatment of pneumococcal infections
 PT such as otitis media, nasopharyngeal and bronchial infections,
 PT comprises Streptococcus pneumoniae proteins
 PS Disclosure; Page 64-65; 70pp; English.
 XX
 CC Although a number of proteins have been suggested as being involved
 CC in the pathogenicity of Streptococcus pneumoniae, there still remains
 CC a need to identify polypeptides having epitopes in common from
 CC various strains of S. pneumoniae in order to utilize such
 CC polypeptides in vaccines to protect against a wide variety of
 CC S. pneumoniae. New vaccine compositions are described which comprise a
 CC Streptococcus pneumoniae polypeptide (or fragments) of 80 - 680 amino
 CC acids in length that comprise at least one histidine triad residue
 CC (HxxHx) or a coiled-coil region, or an antibody directed against
 CC these features. The vaccine is useful in protecting against infection
 CC by Streptococcus pneumoniae. The vaccine composition comprising
 CC antibodies to its useful for passive immunization for treating
 CC pneumococcal infections which includes otitis media, nasopharyngeal
 CC and bronchial infections.
 CC
 XX
 SQ Sequence 2451 BP; 849 A; 467 C; 499 G; 635 T; 1 other;
 Query Match 100.0%; Score 2389; DB 21; Length 2451;
 Best local Similarity 100.0%; Pred. No. 0;
 Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 361 TGCCACGCGGATTAACGTCCTGACCAAAAGAGAAATCATGACCAAAAACAAGACATAG 420
 DB TGCCACGCGGATTAACGTCCTGACCAAAAGAGAAATCATGACCAAAAACAAGACATAG 479
 QY 421 TCACATCGTGAAGGTGAAGTCCCAAGAAACGATGTCGTCTTGGCAGTTCGCA 480
 DB TCACATCGTGAAGGTGAAGTCCCAAGAAACGATGTCGTCTTGGCAGTTCGCA 539
 QY 481 AGACGCTATPCTACAGTGTATGTTATCTTATGCTCTGATATCATAGAGATAC 540
 DB AGACGCTATPCTACAGTGTATGTTATCTTATGCTCTGATATCATAGAGATAC 599
 QY 541 TGTGATGCTTATATGCTTCTCATGAGATCATATACATTACCTTAAGATGAGTT 600
 DB TGTGATGCTTATATGCTTCTCATGAGATCATATACATTACCTTAAGATGAGTT 659
 QY 601 ATCAGCTAGCGAGTGTGCTGTCGAGAAAGCTTCTATCTGTGTGAGAAATCTGTCAA 660
 DB ATCAGCTAGCGAGTGTGCTGTCGAGAAAGCTTCTATCTGTGTGAGAAATCTGTCAA 719
 QY 661 TTCAAGAACCTATGCGCGACAAATATAGCATTAACCTTCAAGAACTGGTACCTTC 720
 DB TTCAAGAACCTATGCGCGACAAATATAGCATTAACCTTCAAGAACTGGTACCTTC 779
 QY 721 TGTAAAGATTCAGAACTACAAATATACACACAGCAACAGCAACACTAACAGTCA 780
 DB TGTAAAGATTCAGAACTACAAATATACACACAGCAACAGCAACACTAACAGTCA 839
 QY 781 AGCAAGTCAAAAGTATATGATATGATATGCTCTTGAACAGCTTACAACTGCTTTGAG 840
 DB AGCAAGTCAAAAGTATATGATATGATATGCTCTTGAACAGCTTACAACTGCTTTGAG 899
 QY 841 TCACAGCATGTAGATCTGATGCTGCTTGTGATTCAGACAAATACAACTGCAAC 900
 DB TCACAGCATGTATGATATGATGCTGCTTGTGATTCAGACAAATACAACTGCAAC 959
 QY 901 AGCTAGAGGTGTTGCACTGCGACACAGAGATCATTAACCTTACCTTACCTCAAT 960
 DB AGCTAGAGGTGTTGCACTGCGACACAGAGATCATTAACCTTACCTTACCTCAAT 1019
 QY 961 GTCTGAATGGAAAGAAAGATGCTGCTGATATATCCCTTGTATGCTTCAACATTTG 1020
 DB GTCTGAATGGAAAGAAAGATGCTGCTGATATATCCCTTGTATGCTTCAACATTTG 1079
 QY 1021 GGTACAGATTCAGAGCGACAGAAACCAAGTCCAAACCGACTCGGAACCTAGTCCAG 1080
 DB GGTACAGATTCAGAGCGACAGAAACCAAGTCCAAACCGACTCGGAACCTAGTCCAG 1139
 QY 1081 CCCGCAACCTGACCAAAATCTTAAATATAGACTCAAAATCTTGTGTTAGTCAAGTGT 1140
 DB CCCGCAACCTGACCAAAATCTTAAATATAGACTCAAAATCTTGTGTTAGTCAAGTGT 1199
 QY 1141 ACGAAAGTGGGGAAGATATGATTCGAAGAAAGGCGATCTCTGTTATGCTTTGTC 1200
 DB ACGAAAGTGGGGAAGATATGATTCGAAGAAAGGCGATCTCTGTTATGCTTTGTC 1259
 QY 1201 GAAAGATTTTCAATCTGAAACCTGTTAAATCTTGAAGCAAGTATCAAAACAAGAG 1260
 DB GAAAGATTTTCAATCTGAAACCTGTTAAATCTTGAAGCAAGTATCAAAACAAGAG 1319
 QY 1261 TGTTCACACACTTAACTGCTAAATAAAAGAAATGTTGCTCTGTCGACCAAGATTTTA 1320
 DB TGTTCACACACTTAACTGCTAAATAAAAGAAATGTTGCTCTGTCGACCAAGATTTTA 1379
 QY 1321 TGAATTAAGCATATATCTGTTAACTGAGGCTCATAAAGCTTGTGTTGAAATTAAGGTCG 1380
 DB TGAATTAAGCATATATCTGTTAACTGAGGCTCATAAAGCTTGTGTTGAAATTAAGGTCG 1439
 QY 1381 TAAATTCGATTTTCAACCTTAAAGCAATATTAAGAAAGCTGTAATGATGAATGCACTAA 1440
 DB TAAATTCGATTTTCAACCTTAAAGCAATATTAAGAAAGCTGTAATGATGAATGCACTAA 1499
 QY 1441 TAAAGAAAAATGATGATGATTTATGGAATTCCTGACCAATTAACCTCAAGACG 1500

```

Db      1500 TAAAGAAAAATGGATGATGATTATTTGGCATTCTAGAGCAACCAATTCACAGAGG 1559
Oy      1501 ACTTGCAAAACCAATTTCTCAATGAGATATCTGAAGCAAGTTGATGCTCAAT 1560
Db      1560 ACTTGCAAAACCAATTTCTCAATGAGATATCTGAAGCAAGTTGATGCTCAAT 1619
Oy      1561 AGCTGATATGATACAGCTGATGTTTACATTTTGAAGCAATGATATATGATGA 1620
Db      1620 AGCTGATATGATACAGCTGATGTTTACATTTTGAAGCAATGATATATGATGA 1679
Oy      1621 TGAAGAGATGATATGATGATGCTGATATGAGGCAATGATGATGATGAAAGATG 1680
Db      1680 TGAAGAGATGATATGATGATGCTGATATGAGGCAATGATGATGATGAAAGATG 1739
Oy      1681 CTTTCTGATAGAAAAAGTTGAGCTCAACCTATTAAGAAAAAGATGCTTACC 1740
Db      1740 CTTTCTGATAGAAAAAGTTGAGCTCAACCTATTAAGAAAAAGATGCTTACC 1799
Oy      1741 TCCATCTCCAGACGAGATGTTAAAGCAAACTCAAGATGATGACAGCTATTTA 1800
Db      1800 TCCATCTCCAGACGAGATGTTAAAGCAAACTCAAGATGATGACAGCTATTTA 1859
Oy      1801 CAATGCTGAAAGGGGAAAAAGCAATTCCTCTGCTGATTCATATATGTTGAGA 1860
Db      1860 CAATGCTGAAAGGGGAAAAAGCAATTCCTCTGCTGATTCATATATGTTGAGA 1919
Oy      1861 TACAGTTAGGTTAAAAACGGTAAATTTGATTTCTCTCAATGATGATTAAT 1920
Db      1920 TACAGTTAGGTTAAAAACGGTAAATTTGATTTCTCTCAATGATGATTAAT 1979
Oy      1921 TAAATTTGCTGTTGATGATCACAACATCAAAAGTCCAAATGGCTATACCTTGAAGA 1980
Db      1980 TAAATTTGCTGTTGATGATCACAACATCAAAAGTCCAAATGGCTATACCTTGAAGA 2039
Oy      1981 TTTGTTGCGAGATTAAGTACTAGTAGAACACCTGCGAAGCTGCAATTTCTAATGA 2040
Db      2040 TTTGTTGCGAGATTAAGTACTAGTAGAACACCTGCGAAGCTGCAATTTCTAATGA 2099
Oy      2041 TGGATGGGCGAATGCGACGATGATGTTAGGCAAGAAAGCCACAGTGAAGATCCAA 2100
Db      2100 TGGATGGGCGAATGCGACGATGATGTTAGGCAAGAAAGCCACAGTGAAGATCCAA 2159
Oy      2101 TAAAGACTTCAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGGTGAAGAGTCCC 2160
Db      2160 TAAAGACTTCAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGGTGAAGAGTCCC 2219
Oy      2161 TCAAGTAGAGCTGAAAAAGTGAAGCCCACTCAAGAGAGAGAGTTTCTTGCGAA 2220
Db      2220 TCAAGTAGAGCTGAAAAAGTGAAGCCCACTCAAGAGAGAGAGTTTCTTGCGAA 2279
Oy      2221 AGTAAAGGATTTAGTCTGAAGCCCAATGCAAGAACTTACTGCTTGAAGATGA 2280
Db      2280 AGTAAAGGATTTAGTCTGAAGCCCAATGCAAGAACTTACTGCTTGAAGATGA 2339
Oy      2281 TTTGACTCTTCAATATGATGATGATGATGATGATGATGATGATGATGATGATG 2340
Db      2340 TTTGACTCTTCAATATGATGATGATGATGATGATGATGATGATGATGATGATG 2399
Oy      2341 GTTGTAAAGAGAGTAACTCTTCACTGTAGTGAAGAAAAATTAAC 2389
Db      2400 GTTGTAAAGAGAGTAACTCTTCACTGTAGTGAAGAAAAATTAAC 2448

```

```

RESULT 4
ABX06886
ID ABX06886 standard; DNA; 2406 BP.
AC ABX06886;
XX
XX
XX 11-FEB-2003 (first entry)
DT
XX
DE S. pneumoniae type 4 strain coding region #1174.

```

```

XX Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media;
KM ear infection; anti-inflammatory; antibacterial; immunostimulant;
KM auditory; respiratory; gene therapy; vaccine.
OS Streptococcus pneumoniae type 4 strain.
XX
XX WO200277021-A2.
XX
XX 03-OCT-2002.
XX
XX 27-MAR-2002; 2002WO-IB02163.
XX
XX 27-MAR-2001; 2001GB-0007658.
XX
XX (CHIR-) CHIRON SPA.
XX (GENO-) INST GENOMIC RES.
XX
XX Masignani V, Tettelin H, Fraser C;
XX
XX WPI: 2003-040579/03.
XX P-PSDB: AB01598.
XX
XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,
XX useful as medicaments for treating or preventing a disease or infection
XX due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
XX or ear infection
XX
XX Claim 6; SEQ ID No 2347; 56bp; English.
XX
XX The invention relates to a protein comprising or having at least 50%
XX identity to any of the 2469 amino acid sequences, identified in the
XX specification (available on a computer readable format), or its fragment,
XX expressed from 2469 or 2489 identified DNA coding regions from the
XX Streptococcus pneumoniae type 4 strain genomic sequence appearing as
XX AB556454. Also included are an antibody which binds one of the
XX proteins, treating a patient by administering the protein, DNA or
XX antibody (in a composition), a kit comprising first and second primers,
XX which are the nucleic acid cited above or fragments between nucleotides
XX 8-100 of a sequence not defined in the specification, for amplifying a
XX target sequence contained within a Streptococcus nucleic acid sequence,
XX where the first primer is substantially complementary to the target
XX sequence and the second primer is substantially complementary to the
XX complement of the target sequence, and where the parts of the primers
XX having substantial complementarity define the termini of the target
XX sequence to be amplified, assay comprising contacting a test compound
XX with the protein, and determining whether the test compound binds to the
XX protein and a Streptococcus pneumoniae bacterium, where one or more
XX genes encoding the proteins has been rendered inactive. The proteins,
XX nucleic acid molecules, antibody and compositions are useful as
XX medicaments for treating or preventing a disease or infection due to
XX streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,
XX sepsis, otitis media or ear infection. They are also useful in developing
XX vaccines, diagnostics and antibiotics. The methods are useful for
XX identifying immunodominant proteins. The present sequence is one of
XX the 2469 identified coding region from the genomic sequence.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 2406 BP; 834 A; 461 C; 490 G; 621 T; 0 other;
XX
XX
XX Query Match 97.9%; Score 2338; DB 25; Length 2406;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1 TTCTTACAGAGTGGAGCTGTATCAAGCTAGAACGTTAAGAAAAATATCGTTTCTTA 60
XX
XX 18 TTCTTACAGAGTGGAGCTGTATCAAGCTAGAACGTTAAGAAAAATATCGTTTCTTA 77
XX
XX 61 TATAGATGAAAAACAAGCGACGCAAAAAAGAGATTTGACTCTGATGAGGTTAGCAA 120
XX

```

Db 78 TATAGATGAAAAACAAGCGAACGAAAAACGAGATTGACTCCTGATGAGTTAGCAA 137
Qy 121 GCGTGAAGGAATCAATGCTGAGCAAAATGTCATCAAGATTAACAGACCAAGGCTATGTCAC 180
Db 136 GCGTGAAGGAATCAATGCTGAGCAAAATGTCATCAAGATTAACAGACCAAGGCTATGTCAC 197
Qy 181 TTCAATGAGGCAACCTATCATTTATTAACATGTAAGGTCCTTATGAACGCTATCATGAG 240
Db 198 TTCAATGAGGCAACCTATCATTTATTAACATGTAAGGTCCTTATGAACGCTATCATGAG 257
Qy 241 TGAAGATTAATCATGAAAGATCCAACTATTAAGTAAAGATGAGATTTGTTAATGA 300
Db 258 TGAAGATTAATCATGAAAGATCCAACTATTAAGTAAAGATGAGATTTGTTAATGA 317
Qy 301 GGTCAAGGGTGGATATGTTATCAAGGTAGTGAATAATGTTATCTTAAAGATGC 360
Db 318 GGTCAAGGGTGGATATGTTATCAAGGTAGTGAATAATGTTATCTTAAAGATGC 377
Qy 361 TGCCCAAGCGGATTAACGTCCTGACAAAAGAGAAATCAATGACAAAACAAAGACATAG 420
Db 378 TGCCCAAGCGGATTAACGTCCTGACAAAAGAGAAATCAATGACAAAACAAAGACATAG 437
Qy 421 TCAACATGTAAGGTGGAATCTCAAGAAAGATGCTGTTGCTTGGCAAGTTGCA 480
Db 438 TCAACATGTAAGGTGGAATCTCAAGAAAGATGCTGTTGCTTGGCAAGTTGCA 497
Qy 481 AGGACGCTATCTACAGATGATGTTATATCTTTAATGCTTGTATCATAGAGGATAC 540
Db 498 AGGACGCTATCTACAGATGATGTTATATCTTTAATGCTTGTATCATAGAGGATAC 557
Qy 541 TGTGATGCTTATATCGTTCCATGAGATCATTAACATTACATTCCTTAAGAAATGAGTT 600
Db 558 TGTGATGCTTATATCGTTCCATGAGATCATTAACATTACATTCCTTAAGAAATGAGTT 617
Qy 601 ATCAAGTACGAGATTTGGCTGCTGACAAAGCCTTCTATCTGCTGAGGAAATCTGTCAA 660
Db 618 ATCAAGTACGAGATTTGGCTGCTGACAAAGCCTTCTATCTGCTGAGGAAATCTGTCAA 677
Qy 661 TTCAAGAACTATGCGCGCAAAAATAGGATTAACCTTCAAGAACTAGGATACCTTC 720
Db 678 TTCAAGAACTATGCGCGCAAAAATAGGATTAACCTTCAAGAACTAGGATACCTTC 737
Qy 721 TGTAAAGCAATCCAGAACTAATAATCTAACAAGCAAAACAGCAACACTTAACAGTCA 780
Db 738 TGTAAAGCAATCCAGAACTAATAATCTAACAAGCAAAACAGCAACACTTAACAGTCA 797
Qy 781 AGCAAGTCAAAAGTAATGATCATTTGATAGTCTCTTGAACAGCTCTAACAATGCTTTGAG 840
Db 798 AGCAAGTCAAAAGTAATGATCATTTGATAGTCTCTTGAACAGCTCTAACAATGCTTTGAG 857
Qy 841 TCAACGACATGATGAATGTAAGGCTGCTGTTGATCCAGCAAAATCACAAGTCGAC 900
Db 858 TCAACGACATGATGAATGTAAGGCTGCTGTTGATCCAGCAAAATCACAAGTCGAC 917
Qy 901 AGCTAGAGGTGTCAGTGCACACAGAGATCATTAACACTTACCTTACTCTCAAT 960
Db 918 AGCTAGAGGTGTCAGTGCACACAGAGATCATTAACACTTACCTTACTCTCAAT 977
Qy 961 GTCTGAATTTGAAGAAAGATGCTGTAATTAATCCCTTGTATGCTTCAAAACATG 1020
Db 978 GTCTGAATTTGAAGAAAGATGCTGTAATTAATCCCTTGTATGCTTCAAAACATG 1037
Qy 1021 GGTACCGAATTCAGAGCCAGAAACCAAGTCCACACCGACTCCGGAACCTAGTCAGG 1080
Db 1038 GGTACCGAATTCAGAGCCAGAAACCAAGTCCACACCGACTCCGGAACCTAGTCAGG 1097
Qy 1081 CCCCACACCTGACCAAAATCTTTAAATAGACTCAAAATCTTCTTTGTTAGTACAGTGT 1140
Db 1098 CCCCACACCTGACCAAAATCTTTAAATAGACTCAAAATCTTCTTTGTTAGTACAGTGT 1157
Qy 1141 ACGAAAGTTGGGAGAGATATGTAATTCGAAAGAAAGGACATCTCTGTTATGCTTTGTC 1200
Db 1158 ACGAAAGTTGGGAGAGATATGTAATTCGAAAGAAAGGACATCTCTGTTATGCTTTGTC 1217

Qy 1201 GAAAGATTTACATCTGAAAATGTTAAATAATCTTGAAGCAAGTTATCAAAACAAGAGAG 1260
Db 1218 GAAAGATTTACATCTGAAAATGTTAAATAATCTTGAAGCAAGTTATCAAAACAAGAGAG 1277
Qy 1261 TGTTCACACCTTTAACTGCTAAAAAAGAAATGTTGCTCTCTGACCAAGAAATTTTA 1320
Db 1278 TGTTCACACCTTTAACTGCTAAAAAAGAAATGTTGCTCTCTGACCAAGAAATTTTA 1337
Qy 1321 TGAATTAAGCATATTAATCTGTTAACTGAGGCTCAATAAGCCTTGTGTTAAATAAGGCTG 1380
Db 1338 TGAATTAAGCATATTAATCTGTTAACTGAGGCTCAATAAGCCTTGTGTTAAATAAGGCTG 1397
Qy 1381 TAAATTCGATTTCCAGCCTTAGACAATAATTAAGAAGCTTGAATGATGAATGACTTAA 1440
Db 1398 TAAATTCGATTTCCAGCCTTAGACAATAATTAAGAAGCTTGAATGATGAATGACTTAA 1457
Qy 1441 TAAAGAAAAATGGTATGATATTTATTTGGCAATTCCTAGCAACAAATTAACCAACAGAG 1500
Db 1458 TAAAGAAAAATGGTATGATATTTATTTGGCAATTCCTAGCAACAAATTAACCAACAGAG 1517
Qy 1501 ACTTGGCAAAACCAATTTCTCAAAATGAGTATCTGAAGAGCAAGTGTGTAATTCCTCAAT 1560
Db 1518 ACTTGGCAAAACCAATTTCTCAAAATGAGTATCTGAAGAGCAAGTGTGTAATTCCTCAAT 1577
Qy 1561 AGCTGATTAATTAATCAAGTCAAGTGGTTACATTTTGTGATGAACATGATTAATCAAGTGA 1620
Db 1578 AGCTGATTAATTAATCAAGTCAAGTGGTTACATTTTGTGATGAACATGATTAATCAAGTGA 1637
Qy 1621 TGAAGAGATGATATATGTAACGCTCATATGAGGCAATGCTCATGATTTGAAAAAGATAG 1680
Db 1638 TGAAGAGATGATATATGTAACGCTCATATGAGGCAATGCTCATGATTTGAAAAAGATAG 1697
Qy 1681 CCTTTCTGATTAAGAAAAAGTTGAGCTCAAGCTTAATCTTAAGAAAAAGATATCTTAAC 1740
Db 1698 CCTTTCTGATTAAGAAAAAGTTGAGCTCAAGCTTAATCTTAAGAAAAAGATATCTTAAC 1757
Qy 1741 TCCATCTCCAGACGACAGATGTTAAAGCAAAATCCAACTGAGATPAGTGCAGAGCTATTTTA 1800
Db 1758 TCCATCTCCAGACGACAGATGTTAAAGCAAAATCCAACTGAGATPAGTGCAGAGCTATTTTA 1817
Qy 1801 CAATCGTGTGAAGGGGAAAAACGAATTCACCTGTTGCACTTCCATATATGTTGAGCA 1860
Db 1818 CAATCGTGTGAAGGGGAAAAACGAATTCACCTGTTGCACTTCCATATATGTTGAGCA 1877
Qy 1861 TACAGTTGAGGTTAAAAACGTAATTTGATTTATCTTCAATTAAGATTAATTAACCTAATAT 1920
Db 1878 TACAGTTGAGGTTAAAAACGTAATTTGATTTATCTTCAATTAAGATTAATTAACCTAATAT 1937
Qy 1921 TAAATTTGCTGTTGTTGATGATCAACATPACAAAGCTCCAAATGCTATACCTTGAAGA 1980
Db 1938 TAAATTTGCTGTTGTTGATGATCAACATPACAAAGCTCCAAATGCTATACCTTGAAGA 1997
Qy 1981 TTTGTTTGCAGCATTAAGTACTACGTAGAACACCTTGAAGAGCTGACCAATTTCTAATGA 2040
Db 1998 TTTGTTTGCAGCATTAAGTACTACGTAGAACACCTTGAAGAGCTGACCAATTTCTAATGA 2057
Qy 2041 TGCATGGGGCAATGCCAGTGAACATGTTTATAGCAAAAGAACCAACAGTGAAGTCCAAA 2100
Db 2058 TGCATGGGGCAATGCCAGTGAACATGTTTATAGCAAAAGAACCAACAGTGAAGTCCAAA 2117
Qy 2101 TAAAGACTTCAAAAGCGATGTAAGAGCCAGTGAAGAAACCTGCTGACGAGAAAGTCCC 2160
Db 2118 TAAAGACTTCAAAAGCGATGTAAGAGCCAGTGAAGAAACCTGCTGACGAGAAAGTCCC 2177
Qy 2161 TCAAGTAGAGACTGAAAAAAGTAGAAGCCCAACTCAAAAGAACAGAAATTTTCTTGGCAA 2220
Db 2178 TCAAGTAGAGACTGAAAAAAGTAGAAGCCCAACTCAAAAGAACAGAAATTTTCTTGGCAA 2237
Qy 2221 AGTAAGGATTTCTATGTTGAAGCAATGCAACGAAATCTTACCTGTTTACGAATTA 2280
Db 2238 AGTAAGGATTTCTATGTTGAAGCAATGCAACGAAATCTTACCTGTTTACGAATTA 2297

QY 2281 TTGACTCTTCAAAATTGATGATACATATCATGACGAGAGCAAAAAATTACTTGC 2340
 Db 2298 TTGACTCTTCAAAATTGATGATACATATCATGACGAGAGCAAAAAATTACTTGC 2357
 QY 2341 GTTGTAAAGAGAGATATCTTCTCATCTGTAGTAAGGAAAAATTAAC 2389
 Db 2358 GTTGTAAAGAGAGATATCTTCTCATCTGTAGTAAGGAAAAATTAAC 2406

RESULT 5
 AAV52227
 ID AAV52227 standard; DNA; 8195 BP.
 AC AAV52227;
 XX
 XX
 DT 23-OCT-1998 (first entry)
 XX
 DE Streptococcus pneumoniae genome fragment SEQ ID NO:94.
 XX
 KM Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
 XX computer readable medium; vaccine; pharmaceutical composition; ds.
 OS Streptococcus pneumoniae.
 XX
 PN WO9818931-A2.
 XX
 PD 07-MAY-1998.
 XX
 PF 30-OCT-1997; 97MO-US19588.
 XX
 PR 31-OCT-1996; 96US-0029960.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
 PI Kunsch CA, Rosen CA;
 XX
 DR WPI; 1998-272225/24.
 XX

Computer-readable medium with recorded Streptococcus pneumoniae polynucleotide sequences - useful in diagnostic kits and assays, and pharmaceutical compositions and vaccines for Streptococcus pneumoniae
 Claim 1; Page 727-732; 1409pp; English.

The present invention describes a computer readable medium which has the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524) recorded on it, or a representative fragment or a sequence at least 95% identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from Streptococcus pneumoniae. The present invention also describes an isolated nucleic acid molecule encoding a homologue of any of the fragments of the S. pneumoniae genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced by a process comprising: (a) screening a genomic DNA library using as a probe a target sequence defined by any of the sequences in SEQ ID NO:1 to 391, identifying members of the library which contain sequences that hybridize to the target sequence and isolating the nucleic acid molecules from the members; or (b) isolating mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid molecules whose nucleotide sequence is homologous to amplification primers derived from the fragment of the S. pneumoniae genome to prime the amplification and isolating the amplified sequences. The computer readable medium can be used in a computer-based system for identifying fragments of the S. pneumoniae genome of commercial importance, or expression modulating fragments of the S. pneumoniae genome. Products from the present invention can be used in diagnosis kits and assays, and pharmaceutical compositions and vaccines for S. pneumoniae.

Sequence 8195 BP; 2688 A; 1622 C; 1777 G; 2105 T; 3 other;

Query Match 97.9%; Score 2338; DB 19; Length 8195;
 Best Local Similarity 100.0%; Pred. No. 0;

Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	TTCTTAAGAGTGGGACGTGATCAAGCTAAGACGGTTAAGGAAATTAATCGTGTCCGA	60
Db	3053	TTCTTAAGAGTGGGACGTGATCAAGCTAAGACGGTTAAGGAAATTAATCGTGTCCGA	3112
QY	61	TATAGATGAAAAACAAAGGACGCAAAAAACGAGAAATTGACTCCTGATAGGTTAGCA	120
Db	3113	TATAGATGAAAAACAAAGGACGCAAAAAACGAGAAATTGACTCCTGATAGGTTAGCA	3172
QY	121	GGGTGAAGGAATCAATGCTGAGCAAAATGTCATCAAGATTAACAGCAAGGCTATGTCAC	180
Db	3173	GGGTGAAGGAATCAATGCTGAGCAAAATGTCATCAAGATTAACAGCAAGGCTATGTCAC	3232
QY	181	TTACATAGGCGACCTATCAATTAATTAAGTGAAGTCTTATGACGCTATCATCAG	240
Db	3233	TTACATAGGCGACCTATCAATTAATTAAGTGAAGTCTTATGACGCTATCATCAG	3292
QY	241	TGAAGATTTACTCATGAAGAATCCAAACTATTAAGCTAAAGATGAGATATTGTTAATGA	300
Db	3293	TGAAGATTTACTCATGAAGAATCCAAACTATTAAGCTAAAGATGAGATATTGTTAATGA	3352
QY	301	GGTCAAGGATGATATGTTATCAAGTATGAGAAATATCTATGTTAAGGATGC	360
Db	3353	GGTCAAGGATGATATGTTATCAAGTATGAGAAATATCTATGTTAAGGATGC	3412
QY	361	TGCCACGCGGATTAACGTCGCTACAAAGAGAAATCAATGACAAAAACAAAGACATAG	420
Db	3413	TGCCACGCGGATTAACGTCGCTACAAAGAGAAATCAATGACAAAAACAAAGACATAG	3472
QY	421	TCAACATGATGAAGGTGAATCTCCAAAGAAAGATGCTGTGCTTGCGACGTTCCGA	480
Db	3473	TCAACATGATGAAGGTGAATCTCCAAAGAAAGATGCTGTGCTTGCGACGTTCCGA	3532
QY	481	AGGACGCTATCTACAGATGATGTTATATCTTTAATGCTTGTGATATCAAGAGATAC	540
Db	3533	AGGACGCTATCTACAGATGATGTTATATCTTTAATGCTTGTGATATCAAGAGATAC	3592
QY	541	TGGTGAATGCTTATATCGTTCCTCATGAGATCATTAATCATTTCTTAAGATGATG	600
Db	3593	TGGTGAATGCTTATATCGTTCCTCATGAGATCATTAATCATTTCTTAAGATGATG	3652
QY	601	ATCAGCTGACGAGTGGCTGCTGCAAGACCTTCTTATCTGTGAGGAAATCTGTCAA	660
Db	3653	ATCAGCTGACGAGTGGCTGCTGCAAGACCTTCTTATCTGTGAGGAAATCTGTCAA	3712
QY	661	TTCAAGAACTATGCGCCGACAAATAGGATTAACATTCAGAAACAACTGGGTACCTTC	720
Db	3713	TTCAAGAACTATGCGCCGACAAATAGGATTAACATTCAGAAACAACTGGGTACCTTC	3772
QY	721	TGTAAAGCATCGAGAACTTAACAATCTTAACAAGCAACAACGCAACTACAGTCA	780
Db	3773	TGTAAAGCATCGAGAACTTAACAATCTTAACAAGCAACAACGCAACTACAGTCA	3832
QY	781	AGCAAGTCAAAAGTATGATGATTAAGTCTCTGTAAGACAGCTCTCAAAAGCTTTGAG	840
Db	3833	AGCAAGTCAAAAGTATGATGATTAAGTCTCTGTAAGACAGCTCTCAAAAGCTTTGAG	3892
QY	841	TCAAGCATGTAGATGTATGATGCTTGTGATGATGATGATGATGATGATGATGATG	900
Db	3893	TCAAGCATGTAGATGTATGATGCTTGTGATGATGATGATGATGATGATGATGATG	3952
QY	901	AGCTGAGAGTGTGAGTGCACACGAGATCATTAACATTCATCCTTACTTCAAT	960
Db	3953	AGCTGAGAGTGTGAGTGCACACGAGATCATTAACATTCATCCTTACTTCAAT	4012
QY	961	GCTGAATTTGAAGAAAGATGCTGCTGATTAATTCCTGTTATGCTTCAACATTTG	1020
Db	4013	GCTGAATTTGAAGAAAGATGCTGCTGATTAATTCCTGTTATGCTTCAACATTTG	4072
QY	1021	GGTACCAAGATTCAGAGCCAGAAACAAAGTCCAAACGCACTCGGAACCTAGTCCAG	1080
Db	4073	GGTACCAAGATTCAGAGCCAGAAACAAAGTCCAAACGCACTCGGAACCTAGTCCAG	4132

QY 1081 CCCGCAACCTGCACCAATCTTAAATAGACTCAATCTTCTTGGTTAGTCAGCTGCT 1140
 Db 4133 CCCGCAACCTGCACCAATCTTAAATAGACTCAATCTTCTTGGTTAGTCAGCTGCT 4192
 QY 1141 ACGAAAGTTGGGGAGGAGATATGTTGGAAGAAAAGGCGATCTCTGTTATGTTCTTGC 1200
 Db 4193 ACGAAAGTTGGGGAGGAGATATGTTGGAAGAAAAGGCGATCTCTGTTATGTTCTTGC 4252
 QY 1201 GAAAGATTACCATCTGAAACGTGTAATAAATCTTGAAGCAAGTATGTAACCAAGAG 1260
 Db 4253 GAAAGATTACCATCTGAAACGTGTAATAAATCTTGAAGCAAGTATGTAACCAAGAG 4312
 QY 1261 TGTTCACACACTTAACTGCTAAAAAAGTCTCTCTCTGTCGACCAAGATTTTA 1320
 Db 4313 TGTTCACACACTTAACTGCTAAAAAAGTCTCTCTCTGTCGACCAAGATTTTA 4372
 QY 1321 TGATTAAGCATTAATCTGTTACTGAGGCTCATTAAGCCTTGTGTAATAAGGCTG 1380
 Db 4373 TGATTAAGCATTAATCTGTTACTGAGGCTCATTAAGCCTTGTGTAATAAGGCTG 4432
 QY 1381 TAAATTCGATTCTCAAGCCTTAGACAAATTAATTAAGCCTTGAATGATGACCTAA 1440
 Db 4433 TAAATTCGATTCTCAAGCCTTAGACAAATTAATTAAGCCTTGAATGATGACCTAA 4492
 QY 1441 TAAAGAAAATTTGGTAGATGATTTATTTGGCAATCTCTAGACCAATTCACAGAG 1500
 Db 4493 TAAAGAAAATTTGGTAGATGATTTATTTGGCAATCTCTAGACCAATTCACAGAG 4552
 QY 1501 ACTTGGCAACCAATTTCTCAATTTGAGTATATCTGAGACGAAAGTTGCTATGCTCAAT 1560
 Db 4553 ACTTGGCAACCAATTTCTCAATTTGAGTATATCTGAGACGAAAGTTGCTATGCTCAAT 4612
 QY 1561 AGCTGATTAAGTATACAGTCAAGTGTATACATTTTATGAACTGATATATACAGTGA 1620
 Db 4613 AGCTGATTAAGTATACAGTCAAGTGTATACATTTTATGAACTGATATATACAGTGA 4672
 QY 1621 TGAAGGAGATGATATGTAAGGCTCATATGAGGCTCATATGCTGATTTGAAAAAGATAG 1680
 Db 4673 TGAAGGAGATGATATGTAAGGCTCATATGAGGCTCATATGCTGATTTGAAAAAGATAG 4732
 QY 1681 CCTTTCTGATTAAGAAAAGTTGAGCTCAAGCCTATATCTAAGAAAAGGATCTTACC 1740
 Db 4733 CCTTTCTGATTAAGAAAAGTTGAGCTCAAGCCTATATCTAAGAAAAGGATCTTACC 4792
 QY 1741 TCCATCTCCAGACGAGATGTTAAAGCAATTCCACTGAGATAGTGCAGACGCTATTTA 1800
 Db 4793 TCCATCTCCAGACGAGATGTTAAAGCAATTCCACTGAGATAGTGCAGACGCTATTTA 4852
 QY 1801 CAATCGTGTGAAGGGGAAAAAGCAATTCGACTGCTGCTGCTCATATATGTTGAGCA 1860
 Db 4853 CAATCGTGTGAAGGGGAAAAAGCAATTCGACTGCTGCTGCTCATATATGTTGAGCA 4912
 QY 1861 TACAGTTGAGGTTAAAAAGGTAATTTGATTTCTCATTAAGATCAATTAATAT 1920
 Db 4913 TACAGTTGAGGTTAAAAAGGTAATTTGATTTCTCATTAAGATCAATTAATAT 4972
 QY 1921 TAAATTTGCTTGGTTGATGATCAGACATACAAAGCTCCAAATGGCTTACTTGGAGA 1980
 Db 4973 TAAATTTGCTTGGTTGATGATCAGACATACAAAGCTCCAAATGGCTTACTTGGAGA 5032
 QY 1981 TTTGTTTGGCGAGATTAGTACTAGTGAACACCTGACGAAAGTCCAAATTCGAATGA 2040
 Db 5033 TTTGTTTGGCGAGATTAGTACTAGTGAACACCTGACGAAAGTCCAAATTCGAATGA 5092
 QY 2041 TGAATGGGCGAATGCGACGTAGACATGTTAGGCAAGAAAGCAACAGTGAAGATCCAA 2100
 Db 5093 TGAATGGGCGAATGCGACGTAGACATGTTAGGCAAGAAAGCAACAGTGAAGATCCAA 5152
 QY 2101 TAAAGACTTCAAAAGCGAGTGAAGAGCCAGTGAAGAAACACTGTGAGCCAGAAAGTCCC 2160
 Db 5153 TAAAGACTTCAAAAGCGAGTGAAGAGCCAGTGAAGAAACACTGTGAGCCAGAAAGTCCC 5212

QY 2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAGAAAGCAGAGTTTGTCTGGCAA 2220
 Db 5213 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAGAAAGCAGAGTTTGTCTGGCAA 5272
 QY 2221 AGTAAAGGATTTCTAGTCTGAAAGCCAAATGCAACGAAATCTTACTGTTTGAATTA 2280
 Db 5273 AGTAAAGGATTTCTAGTCTGAAAGCCAAATGCAACGAAATCTTACTGTTTGAATTA 5332
 QY 2281 TTTGACTCTTCAATTTATGATTAACAATAGTATGATGAGCAAGCAAGAAATTTACTTGC 2340
 Db 5333 TTTGACTCTTCAATTTATGATTAACAATAGTATGATGAGCAAGCAAGAAATTTACTTGC 5392
 QY 2341 GTTGTAAAAAGAGTAAATCTTCAATCTGTAAGTAAAGAAAAATTAAC 2389
 Db 5393 GTTGTAAAAAGAGTAAATCTTCAATCTGTAAGTAAAGAAAAATTAAC 5441

RESULT 6
 ABS56454/C
 ID ABS56454 standard; DNA; 2162598 BP.
 XX
 AC ABS56454;
 XX
 DT 10-FEB-2003 (first entry)
 XX
 DE Streptococcus pneumoniae type 4 strain complete genome.
 XX
 KM ds; bacterial meningitis; pneumonia; sepsis; otitis media; genome;
 KM ear infection; antiinflammatory; antibacterial; immunostimulant;
 KM auditory; respiratory; gene therapy; vaccine.
 OS Streptococcus pneumoniae type 4 strain.
 XX
 PN W0200277021-A2.
 XX
 EN 03-OCT-2002.
 PD 27-MAR-2002; 2002WO-1B02163.
 PF 27-MAR-2001; 2001GB-0007658.
 PR 27-MAR-2001; 2001GB-0007658.
 XX
 PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Masignani V, Tettelein H, Fraser C;
 DR WPI; 2003-040579/03.
 XX
 PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,
 PT useful as medicaments for treating or preventing a disease or infection
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
 PT or ear infection -
 XX
 PS Claim 17; SEQ ID No 4979; 56pp; English.
 XX
 CC The invention relates to a protein comprising or having at least 50%
 CC identity to any of the 2469 amino acid sequences, identified in the
 CC specification (available on a computer readable format), or its fragment,
 CC expressed from 2469 of 2489 identified DNA coding regions from the
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
 CC ABS56454. Also included are an antibody which binds one of the
 CC proteins, treating a patient by administering the protein, DNA or
 CC antibody (in a composition), a kit comprising first and second primers,
 CC which are the nucleic acid cited above or fragments between nucleotides
 CC 8-100 of a sequence not defined in the specification, for amplifying a
 CC target sequence contained within a Streptococcus nucleic acid sequence,
 CC where the first primer is substantially complementary to the target
 CC sequence and the second primer is substantially complementary to the
 CC complement of the target sequence, and where the parts of the primers
 CC having substantial complementarity define the termini of the target
 CC sequence to be amplified, assay comprising contacting a test compound
 CC with the protein, and determining whether the test compound binds to the
 CC protein and a Streptococcus pneumoniae bacterium, where one or more

CC genes encoding the proteins has been rendered inactive. The proteins,
CC nucleic acid molecules, antibody and compositions are useful as
CC medicaments for treating or preventing a disease or infection due to
CC streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,
CC sepsis, otitis media or ear infection. They are also useful in developing
CC vaccines, diagnostics and antibiotics. The methods are useful for
CC identifying immunodominant proteins. The present sequence is the
CC Streptococcus pneumoniae type 4 strain genome sequence.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIP0 at
CC ftp.wipo.int/pub/published_poc_sequences.

XX Sequence 2162598 BP; 654373 A; 427176 C; 431369 G; 649680 T; 0 other;

Query Match 97.9%; Score 2338; DB 25; Length 2162598;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTCTTACGAGTTGGAGCTGTATCAAGCTTGAACGGTTAAGAAATATCGTGTCTTA 60
DB 1006950 TTCTTACGAGTTGGAGCTGTATCAAGCTTGAACGGTTAAGAAATATCGTGTCTTA 1006891
QY 61 TATPAGTGAACAAACGACGCGAAGAAATTTGACTCTCGATGAGGTTAGCAA 120
DB 1006890 TATPAGTGAACAAACGACGCGAAGAAATTTGACTCTCGATGAGGTTAGCAA 1006831
QY 121 GGGTGAAGGAATCAATGCTGAGCAATCGTCATCAAGATTAACAGACCAAGGCTATGTCAC 180
DB 1006830 GGGTGAAGGAATCAATGCTGAGCAATCGTCATCAAGATTAACAGACCAAGGCTATGTCAC 1006771
QY 181 TTCACATGCGACCACTATCATTTATTAACATGTAAGTGTCTTATGACGCTATCATGAC 240
DB 1006770 TTCACATGCGACCACTATCATTTATTAACATGTAAGTGTCTTATGACGCTATCATGAC 1006711
QY 241 TGAAGATTTACTCATGAAAGATCCAAACTATTAAGCTTAAAGATGAGATATTGTTATGA 300
DB 1006710 TGAAGATTTACTCATGAAAGATCCAAACTATTAAGCTTAAAGATGAGATATTGTTATGA 1006651
QY 301 GGTCAAGGTTGATATGTTTATCAAGGTAGTGAAGAAATCTATGTTTACTTAAAGATGC 360
DB 1006650 GGTCAAGGTTGATATGTTTATCAAGGTAGTGAAGAAATCTATGTTTACTTAAAGATGC 1006591
QY 361 TSCCCACGCGGATTAACGCTCCGTACAAAGAGAAATCAATCGACAAAGAGAGATAG 420
DB 1006590 TSCCCACGCGGATTAACGCTCCGTACAAAGAGAAATCAATCGACAAAGAGAGATAG 1006531
QY 421 TCAACATGCTGAAGGTGAAGCTTCAAGAAACGATGCTGTGCTTGGCAAGTTGCGA 480
DB 1006530 TCAACATGCTGAAGGTGAAGCTTCAAGAAACGATGCTGTGCTTGGCAAGTTGCGA 1006471
QY 481 AGGAGGCTTACTACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 540
DB 1006470 AGGAGGCTTACTACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 1006411
QY 541 TGGTATGCTTATATGCTTCTCATGAGATATTAACATTCATTCCTTAAGATAGTT 600
DB 1006410 TGGTATGCTTATATGCTTCTCATGAGATATTAACATTCATTCCTTAAGATAGTT 1006351
QY 601 ATCAGCTAGCAGATTGGCTGTGTCAGAAAGCTTCTCATCTGTGTCAGAAATCTGTCAAA 660
DB 1006350 ATCAGCTAGCAGATTGGCTGTGTCAGAAAGCTTCTCATCTGTGTCAGAAATCTGTCAAA 1006291
QY 661 TTCAGAAACTATCGCGGCAAAATATGCGATTAACCTTCAAGAAACAACTGGGTACTTTC 720
DB 1006290 TTCAGAAACTATCGCGGCAAAATATGCGATTAACCTTCAAGAAACAACTGGGTACTTTC 1006231
QY 721 TGTAAAGCAATCCAGGAATACAAATCTATAACAAGCAACAGAGAACTTAACAGTCA 780
DB 1006230 TGTAAAGCAATCCAGGAATACAAATCTATAACAAGCAACAGAGAACTTAACAGTCA 1006171
QY 781 ACCAAGTCAAGTAATGACATTTAGTCTCTTGAAGACAGCTTCAAACTGCTTTGAG 840
DB 1006170 ACCAAGTCAAGTAATGACATTTAGTCTCTTGAAGACAGCTTCAAACTGCTTTGAG 1006110

DB 1006110 ACCAAGTCAAGTAATGACATTTAGTCTCTTGAAGACAGCTTCAAACTGCTTTGAG 1006411
QY 841 TCAAGGACATGTAAGATCTGATGCGCTCTCTTATGATCCAGACAAATACAGTGCAGC 900
DB 1006110 TCAAGGACATGTAAGATCTGATGCGCTCTCTTATGATCCAGACAAATACAGTGCAGC 1006051
QY 901 ACTTGAAGTGTGAGAGTGCACACAGGATCATTAACCACTTATCCCTTACTCTCAAT 960
DB 1006050 ACTTGAAGTGTGAGAGTGCACACAGGATCATTAACCACTTATCCCTTACTCTCAAT 1005991
QY 961 GTCTGAATTTGAAGAACGATGCTGTATTAATTCCTTGTGTTATGCTTCAACCATTC 1020
DB 1005990 GTCTGAATTTGAAGAACGATGCTGTATTAATTCCTTGTGTTATGCTTCAACCATTC 1005931
QY 1021 GGTACCAATTTCAAGCCAGAACCAAGTCCACACCGATCCCGGAACCTATGTCAG 1080
DB 1005930 GGTACCAATTTCAAGCCAGAACCAAGTCCACACCGATCCCGGAACCTATGTCAG 1005871
QY 1081 CCCGCAACCTGACCAAAATCTTAATAATGACTCAAAATCTTCTTGTGTTAGTCAAGTGT 1140
DB 1005870 CCCGCAACCTGACCAAAATCTTAATAATGACTCAAAATCTTCTTGTGTTAGTCAAGTGT 1005811
QY 1141 ACGAAAGTTGGGGAAGATATGATTCGAAGAAAGGCGATCTCTGTTATGCTTTGCG 1200
DB 1005810 ACGAAAGTTGGGGAAGATATGATTCGAAGAAAGGCGATCTCTGTTATGCTTTGCG 1005751
QY 1201 GAAAGATTTACATCTGAAACCTGTTAAATCTTGAAGCAAGTTATCAAAACAAGAG 1260
DB 1005750 GAAAGATTTACATCTGAAACCTGTTAAATCTTGAAGCAAGTTATCAAAACAAGAG 1005691
QY 1261 TGTTCACACACTTAACTGCTTAAAGAAAGAAATGTCCTGCTGACCAAGATTTTA 1320
DB 1005690 TGTTCACACACTTAACTGCTTAAAGAAAGAAATGTCCTGCTGACCAAGATTTTA 1005631
QY 1321 TGATTAAGCATTAATCTGTTAACTGAGGCTCATTAAGGCTTGTGTTAAATTAAGGCTG 1380
DB 1005630 TGATTAAGCATTAATCTGTTAACTGAGGCTCATTAAGGCTTGTGTTAAATTAAGGCTG 1005571
QY 1381 TAAATTCGATTTCCAGGCTTGAACAATTAATTAAGAGCTTGAATGATATCACTTA 1440
DB 1005570 TAAATTCGATTTCCAGGCTTGAACAATTAATTAAGAGCTTGAATGATATCACTTA 1005511
QY 1441 TAAAGAAATTTGTAAGATGATTTATGTCGATTCAGACCAATTAACCAATCCAGAGG 1500
DB 1005510 TAAAGAAATTTGTAAGATGATTTATGTCGATTCAGACCAATTAACCAATCCAGAGG 1005451
QY 1501 ACTTGGCAAAACCAATTTCTCAATTAAGTATTAAGAGCAAGTTCGATTTGCTCAAT 1560
DB 1005450 ACTTGGCAAAACCAATTTCTCAATTAAGTATTAAGAGCAAGTTCGATTTGCTCAAT 1005391
QY 1561 AGCTGATTAATTAACAGCTCAGATGTTTACATTTTGTGTAACATGATATATCACTGA 1620
DB 1005390 AGCTGATTAATTAACAGCTCAGATGTTTACATTTTGTGTAACATGATATATCACTGA 1005331
QY 1621 TGAAGGAGTGAATGTAAGGCTCATATGAGGCAATGCTCAGATTTGGAAGATAG 1680
DB 1005330 TGAAGGAGTGAATGTAAGGCTCATATGAGGCAATGCTCAGATTTGGAAGATAG 1005271
QY 1681 CTTTCTGATTAAGAAAGTTGTCAGCTCAAGGCTTATCTAAAGAAAGATGATCTTACC 1740
DB 1005270 CTTTCTGATTAAGAAAGTTGTCAGCTCAAGGCTTATCTAAAGAAAGATGATCTTACC 1005211
QY 1741 TCCATCTCAGACGAGATGTTAAAGCAATCCAACTGGAAGTATGTCAGAGCTATTATTA 1800
DB 1005210 TCCATCTCAGACGAGATGTTAAAGCAATCCAACTGGAAGTATGTCAGAGCTATTATTA 1005151
QY 1801 CAATCGTGAAGGAGGAGAAAGCAATTCACCTGTTGAGACTTCAATATATGTTGAGCA 1860
DB 1005150 CAATCGTGAAGGAGGAGAAAGCAATTCACCTGTTGAGACTTCAATATATGTTGAGCA 1005091
QY 1861 TACAGTTAGGTTAAAGACGATATTTGATTAATTCCTATTAAGATCAATTAACATTAAT 1920
DB 1005090 TACAGTTAGGTTAAAGACGATATTTGATTAATTCCTATTAAGATCAATTAACATTAAT 1005031

Oy	1921	TAATATTCCTGGTTGATGATCAACAATCAAAACCTCAATATGCTTACTACCTTGGAA	1980
Db	1005030	TAAATTCCTGGTTGATGATCAACAATCAAAACCTCAATATGCTTACTACCTTGGAA	1004971
Oy	1981	TTTGGTTCCGACGATTAAAGTACTACGTAGAACACCCCTGACGACGTCACATTCATATGA	2040
Db	1004970	TTTGGTTCCGACGATTAAAGTACTACGTAGAACACCCCTGACGACGTCACATTCATATGA	1004911
Oy	2041	TGGATGGGGCAATGCGCAGTGCATGTGTATTGGCAAGAAAGCACAATGTAAAGTCCAT	2100
Db	1004910	TGGATGGGGCAATGCGCAGTGCATGTGTATTGGCAAGAAAGCACAATGTAAAGTCCAT	1004851
Oy	2101	TAAGAACTTCAAAAGCGGATGAAGAGCCAGTAGAGGAAAACACTGCTGAGCCAGAAAGTCC	2160
Db	1004850	TAAGAACTTCAAAAGCGGATGAAGAGCCAGTAGAGGAAAACACTGCTGAGCCAGAAAGTCC	1004791
Oy	2161	TCAAGTAGAGACTGAAAAAGTAGAAAGCCCAACTCAAAGAGCAGAGTTTGTCTTGGCA	2220
Db	1004790	TCAAGTAGAGACTGAAAAAGTAGAAAGCCCAACTCAAAGAGCAGAGTTTGTCTTGGCA	1004731
Oy	2221	AGTAACGATTCCTAGCTGGAAGCCCAATGCAACAGAAACTAGCTGCTTACGAAATPA	2280
Db	1004730	AGTAACGATTCCTAGCTGGAAGCCCAATGCAACAGAAACTAGCTGCTTACGAAATPA	1004671
Oy	2281	TTTGACTCTCAAAATTATGATTAACAATAGTATCAATGGCAGAGCAGAAAAATTACTTG	2340
Db	1004670	TTTGACTCTCAAAATTATGATTAACAATAGTATCAATGGCAGAGCAGAAAAATTACTTG	1004611
Oy	2341	GTTGTTAAAGAGAGTATCCTCTCATCTGTAAAGTAAAGAAAAATTAAAC	2389
Db	1004610	GTTGTTAAAGAGAGTATCCTCTCATCTGTAAAGTAAAGAAAAATTAAAC	1004562

CC	RESULT 7
XX	ID AAA65731
AC	AAA65731 standard; DNA; 2523 BP.
XX	AAA65731;
DT	21-NOV-2000 (first entry)
DE	Streptococcus pneumoniae BVH-11 gene SEQ ID NO:3.
XX	
KM	Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KM	propylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KM	otitis media; pneumonia; immunisation; bactericidal; ds.
OS	Streptococcus pneumoniae.
XX	
PN	WO200039299-A2.
PD	06-JUL-2000.
PF	20-DEC-1999; 99WO-CA01218.
PR	23-DEC-1998; 98US-O113800.
PA	(BIOC-) BIOCHEM PHARMA INC.
PI	Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
DR	WPI; 2000-452397/39.
DR	P-PSTB; AAB12716.
PT	Streptococcal antigens useful for vaccinating against e.g. meningiti-
XX	otitis media, bacteraemia and/or pneumonia -
PS	Example 2; Fig 3; 106pp; English.
XX	

The present invention describes nucleic acids (I) encoding protein antigens (II) from Streptococcus pneumoniae. The protein antigens have bactericidal activity. The nucleic acids, encoding the protein

CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g., meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence encodes the
CC *S. pneumoniae* BVH-11 protein antigen.
XX
XX Sequence 2523 BP; 879 A; 523 C; 526 G; 595 T; 0 other;

Query Match	6.2%	Score 148;	DB 21;	Length 2523;
Best Local Similarity	100.0%	Pred. No. 2.3e-64;		
Matches 148;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	699	CAGAAACAACTGGGACCTTGTGAACCAATCCAGAACTCAAAATCAACACAAAGCA	758
Db	758	CAGAAACAACTGGGACCTTGTGAACCAATCCAGAACTCAAAATCAACACAAAGCA	817
Qy	759	ACACAGCAACACTAAACAGTCAAGCAATCAAAATATGACATTGATAGCTCTGTAAC	818
Db	818	ACACAGCAACACTAAACAGTCAAGCAATCAAAATATGACATTGATAGCTCTGTAAC	877
Qy	819	AGCTTACAAACTGCTTTGAGTCAAG	846
Db	878	AGCTTACAAACTGCTTTGAGTCAAG	905

```

RESULT 8
AAA65736
ID      AAA65736 standard; DNA; 2647 BP.
XX
XX      AAA65736;
.AC
XX
XX      21-NOV-2000 (first entry)
DT

```

DE	Streptococcus pneumoniae BVH-11 gene SEQ ID NO:12.
XX	
XX	Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW	prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW	otitis media; pneumonia; immunisation; bactericidal; ds.
XX	
OS	Streptococcus pneumoniae.
XX	
PN	WO200039299-A2.
XX	
PD	06-JUL-2000.
XX	
PF	20-DEC-1999; 99WO-CA01218.
XX	
PR	23-DEC-1998; 98US-0113800.
XX	
PA	(BIOC-) BIOCHEM PHARMA INC.
PI	Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX	
DR	WPI; 2000-452397/39.
XX	
PT	Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT	otitis media, bacteraemia and/or pneumonia -
XX	
PS	
XX	
XX	Example 6; Fig 15; 106pp; English.
CC	The present invention describes nucleic acids (I) encoding protein
CC	antigens (II) from Streptococcus pneumoniae. The protein antigens
CC	have bactericidal activity. The nucleic acids, encoding the proteins
CC	antigens, may be used for the recombinant production of the proteins
CC	they encode. The protein antigens may then be used as vaccines for the
CC	prevention and treatment of Streptococcal infections in mammals
CC	(especially humans) which result in, e.g. meningitis, otitis media,
CC	bacteraemia and/or pneumonia. The present sequence encodes the
CC	S. pneumoniae BVH-11 protein antigen.
CC	
XX	
Sequence	2647 BP; 934 A; 538 C; 556 G; 619 T; 0 other;

CC proteins can be used to inhibit *S. pneumoniae*-mediated C3 degradation.
CC C3-mediated inflammation and rejection in xenotransplantation can be
CC inhibited by expressing the nucleic acid sequences on the surface of an
CC organ of an animal. In particular, the polypeptides are useful for
CC stimulating the immune system and are effective to immunize or treat a
CC mammalian subject against *Streptococcus pneumoniae* infection or
CC colonization.
CC
XX
SQ Sequence 504 BP; 153 A; 125 C; 98 G; 128 T; 0 other;

Query Match 2.2%; Score 53; DB 21; Length 504;
Best Local Similarity 100.0%; Pred. No. 3.7e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 983 GCTGTTATATCCCTTCGTTATCGTTCAACCATGGGTACGATTCAAG 1035
DB 184 GCTGTTATATTCCTTCGTTATCGTTCAACCATGGGTACGATTCAAG 236

RESULT 11
ID AAY25394 standard; DNA; 2163 BP.
XX
AC AAY25394;
XX
DT 19-JUL-1999 (first entry)
XX
DE *Streptococcus pneumoniae* complement C3-degrading protease DNA.
XX
KW Human complement C3-degrading protease; vaccine; infection;
KW meningitis; pneumonia; xerotransplantation; transplant rejection;
KW inflammation; ds.
XX
OS *Streptococcus pneumoniae*.
XX
PN WO915675-A1.
XX
PD 01-APR-1999.
XX
PE 24-SEP-1998; 98MO-US20186.
XX
PR 24-SEP-1997; 97US-0059907.
XX
PA (AMCY) AMERICAN CYANAMID CO.
XX (MINU) UNIV MINNESOTA.
XX
PI Cheng Q, Finkel DJ, Green BA, Hostetter MK, Masi AW;
XX WPI; 1999-254719/21.
XX DR P-PSDB; AAY05753.
XX
PT New isolated human complement C3-degrading proteinase
XX
PS Claim 54; Page 52-54; 66pp; English.
XX
XX This DNA sequence encodes a 79 kDa protein (see AAY05753) of
CC *Streptococcus pneumoniae* serotype 4 that is capable of degrading
CC human complement protein C3 (HCP3). It was identified in the *S.*
CC *pneumoniae* serotype 4 genome by alignment to another novel open
CC reading frame (see AAY25393) that codes for a 20 kDa HCP3 protease
CC (AAY05752) of *S. pneumoniae* serotype 3. This suggested the open
CC reading frame that codes for the 20 kDa protein may be part of a
CC larger open reading frame. Amino acids 1-58 and 90-132 of the 20
CC kDa protein have substantial sequence identity with amino acids
CC 170-227 and 258-300 of the 79 kDa protein. Proteins and peptides or
CC polypeptides containing these regions, and DNA sequences encoding
CC them (nucleotides 507-661 and 827-999 of the present sequence) are
CC claimed. HCP3 proteases and polypeptides can be used as immune
CC system stimulating compositions (claimed). They can produce an
CC immune response against *S. pneumoniae* to immunize or treat a
CC mammalian subject against infection or colonization (claimed).
CC They can produce a B cell response, a T cell response, an
CC epithelial cell response, or an endothelial cell response

CC (claimed). The expression of the proteins on the surface of an
CC organ of an animal used in xenotransplantation can be used to
CC inhibit C3-mediated inflammation and rejection.
XX
SQ Sequence 2163 BP; 737 A; 450 C; 462 G; 514 T; 0 other;

Query Match 2.2%; Score 53; DB 20; Length 2163;
Best Local Similarity 100.0%; Pred. No. 3.9e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 983 GCTGTTATATCCCTTCGTTATCGTTCAACCATGGGTACGATTCAAG 1035
DB 688 GCTGTTATATTCCTTCGTTATCGTTCAACCATGGGTACGATTCAAG 740

RESULT 12
ID AAV27356 standard; DNA; 2290 BP.
XX
AC AAV27356;
XX
DT 02-OCT-1998 (first entry)
XX
DE *Streptococcus pneumoniae* SP0042 nucleotide.
XX
KW *Streptococcus pneumoniae*; antigen; vaccine; infection; diagnosis;
KW detection; pneumonia; otitis media; meningitis; ss.
XX
OS *Streptococcus pneumoniae*.
XX
FH Key Location/Qualifiers
FT CDS 2..2290
FT FT
FT /*tag= a
FT /product= "Sp0042"
FT /transl_except= (pos:152..154,aa:Xaa)
FT /transl_except= (pos:1406..1408,aa:Xaa)
FT /transl_except= (pos:1430..1432,aa:Xaa)
FT FT /note= "no stop codon given; Xaa is unspecified"
XX
PN WO9816930-A2.
XX
PD 07-MAY-1998.
XX
PE 30-OCT-1997; 97WO-US19422.
XX
PR 31-OCT-1996; 96US-0029960.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Choi GH, Hromockyj A, Johnson LS, Kunsch CA;
XX WPI; 1998-272224/24.
XX DR P-PSDB; AAW55095.
XX
PT Nucleic acid encoding antigenic peptide(s) from *Streptococcus*
XX *pneumoniae* - or their epitope-containing fragments, useful in
XX protective or therapeutic vaccines, and for diagnosis
XX
PS Claim 1; Page 61-62; 118pp; English.
XX
XX The present sequence encodes a protein from *Streptococcus pneumoniae*.
CC The nucleic acid sequence encoding the *Streptococcus pneumoniae* protein
CC can be useful in vaccines for inducing protective antibodies against
CC *Streptococcus pneumoniae*, for treatment or prevention of infection e.g.
CC *pneumonia*, otitis media or meningitis. Probes based on the nucleic acid
CC are used to detect *Streptococcus pneumoniae* infection (by usual hybridization or
CC amplification methods), also for isolating *Streptococcus* genes or their
CC allelic variants. The protein can be used similarly to detect specific
CC antibodies in standard immunoassays, especially for diagnosing or
CC monitoring infections. Antibodies which bind the protein are used to
CC detect corresponding antigens, to purify the protein and for passive
CC immunisation (optionally coupled to a toxin). Vaccines are administered,
CC e.g. by injection, orally or through the skin, typically at 0.01-1000

CC (especially 10-300) mu g/ml per dose.

XX Sequence 2290 BP; 766 A; 474 C; 498 G; 547 T; 5 other;

Query Match 2.2%; Score 53; DB 19; Length 2290;
Best Local Similarity 100.0%; Pred. No. 3.9e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 983 GCTCGATTATATCCCTCGTTATCGTTCAACCATTTGGGTACGATTCAAG 1035
DB 944 GCTCGATTATATCCCTCGTTATCGTTCAACCATTTGGGTACGATTCAAG 996

RESULT 13

AB084824
ID AB084824 standard; DNA; 2290 BP.

AC AB084824;

DT 04-SEP-2002 (first entry)

DE S. pneumoniae SP042 nucleotide sequence SEQ ID NO:65.

XX Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
KW antibacterial; Streptococcal infection; detection; gene; ds.

XX Streptococcus pneumoniae.

OS US2002061545-A1.

PN 23-MAY-2002.

PD 22-JAN-2001; 2001US-0765272.

PF 30-OCT-1997; 97US-0961083.

PR (CHOI/) CHOI G H.

PA (KUNS/) KUNSCH C A.

PA (BARA/) BARASH S C.

PA (DILL/) DILLON P J.

PA (DOUG/) DOUGHERTY B.

PA (FANN/) FANNON M R.

PA (ROSE/) ROSEN C A.

XX Chai GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;

PI Rosen CA;

PI WPI; 2002-479261/51.

DR P-PSDB; ABP54589.

XX New Streptococcus pneumoniae antigens, useful for detecting

PT Streptococcus and for preventing or attenuating disease caused by

PT Streptococcus infection -

XX Claim 1; Page 28-29; 70pp; English.

PS AB084792 to AB084804 represents nucleic acid which encode the

CC Streptococcus pneumoniae antigens given in ABP5457 to ABP5469.

CC The S. pneumoniae antigens have antibacterial activity and can be

CC used in vaccines. The S. pneumoniae antigens can also be used to

CC prevent or attenuate a Streptococcal infection in an animal. The

CC polynucleotides encoding the S. pneumoniae antigens can be used to

CC detect Streptococcus nucleic acids. AB084805 to AB08510 represent

CC primers used in the cloning of S. pneumoniae ORFs (open reading frames)

CC which are used in an example from the present invention.

XX Sequence 2290 BP; 766 A; 474 C; 498 G; 547 T; 5 other;

XX Query Match 2.2%; Score 53; DB 24; Length 2290;

XX Best Local Similarity 100.0%; Pred. No. 3.9e-16;

XX Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 983 GCTCGATTATATCCCTCGTTATCGTTCAACCATTTGGGTACGATTCAAG 1035

DB

944 GCTCGATTATATCCCTCGTTATCGTTCAACCATTTGGGTACGATTCAAG 996

RESULT 14

AAV52376
ID AAV52376 standard; DNA; 2359 BP.

AC AAV52376;

DT 23-OCT-1998 (first entry)

DE Streptococcus pneumoniae genome fragment SEQ ID NO:243.

XX Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;

KW computer readable medium; vaccine; pharmaceutical composition; ds.

XX Streptococcus pneumoniae.

OS -WO9818931-A2.

PN 07-MAY-1998.

PD 30-OCT-1997; 97WO-US19588.

PF 31-OCT-1996; 96US-0029960.

PR (HUMA-) HUMAN GENOME SCI INC.

PA Barash SC, Chai GH, Dillon PJ, Dougherty BA, Fannon M;

PI Kunsch CA, Rosen CA;

PI WPI; 1998-272225/24.

PT Computer-readable medium with recorded Streptococcus pneumoniae

PT polynucleotide sequences - useful in diagnostic kits and assays, and

PT pharmaceutical compositions and vaccines for Streptococcus

PT pneumoniae

PS Claim 1; Page 1265-1266; 1409pp; English.

XX The present invention describes a computer readable medium which has

CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)

CC recorded on it, or a representative fragment or a sequence at least 95%

CC identical to SEQ ID NO:1 to 391. The nucleotide sequences depicted in

CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from

CC Streptococcus pneumoniae. The present invention also describes an

CC isolated nucleic acid molecule encoding a homologue of any of the

CC fragments of the S. pneumoniae genome (SEQ ID NO:1 to 391) where the

CC nucleic acid molecule is produced by a process comprising: (a) screening

CC a genomic DNA library using as a probe a target sequence defined by any

CC of the sequences in SEQ ID NO:1 to 391, identifying members of the

CC library which contain sequences that hybridize to the target sequence and

CC isolating the nucleic acid molecules from the members; or (b) isolating

CC cDNA or RNA produced from an organism, amplifying nucleic acid

CC molecules whose nucleotide sequence is homologous to amplification

CC primers derived from the fragment of the S. pneumoniae genome to prime

CC the amplification and isolating the amplified sequences. The computer

CC readable medium can be used in a computer-based system for identifying

CC fragments of the S. pneumoniae genome of commercial importance, or

CC expression modulating fragments of the S. pneumoniae genome. Products

CC from the present invention can be used in diagnosis kits and assays, and

CC pharmaceutical compositions and vaccines for S. pneumoniae.

XX Sequence 2359 BP; 786 A; 451 C; 511 G; 611 T; 0 other;

XX Query Match 2.2%; Score 53; DB 19; Length 2359;

XX Best Local Similarity 100.0%; Pred. No. 3.9e-16;

XX Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 983 GCTCGATTATATCCCTCGTTATCGTTCAACCATTTGGGTACGATTCAAG 1035

DB 1879 GCTCGATTATATCCCTCGTTATCGTTCAACCATTTGGGTACGATTCAAG 1931

RESULT 15
 ID ABX06885 standard; DNA; 2457 BP.
 AC ABX06885;
 DT 11-FEB-2003 (first entry)
 XX
 DE S. pneumoniae type 4 strain coding region #1173.
 XX
 KW Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media;
 KM ear infection; antiinflammatory; antibacterial; immunostimulant;
 XX auditory; respiratory; gene therapy; vaccine.
 OS Streptococcus pneumoniae type 4 strain.
 XX
 PN WO200277021-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 27-MAR-2002; 2002WO-1B02163.
 XX
 PR 27-MAR-2001; 2001GB-0007658.
 XX
 PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Masignani V, Tettelin H, Fraser C;
 XX
 DR WPI; 2003-040579/03.
 XX
 DR P-PSDB; ABU01597.
 XX
 PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,
 PT useful as medicaments for treating or preventing a disease or infection
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
 PT or ear infection -
 XX
 PS Claim 6; SEQ ID No 2345; 56bp; English.
 XX
 CC The invention relates to a protein comprising or having at least 50%
 CC identity to any of the 2469 amino acid sequences, identified in the
 CC specification (available on a computer readable format), or its fragment,
 CC expressed from 2469 of 2489 identified DNA coding regions from the
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
 CC ABS66454. Also included are an antibody which binds one of the
 CC proteins, treating a patient by administering the protein, DNA or
 CC antibody (in a composition), a kit comprising first and second primers,
 CC which are the nucleic acid cited above or fragments between nucleotides
 CC 8-100 of a sequence not defined in the specification, for amplifying a
 CC target sequence contained within a Streptococcus nucleic acid sequence,
 CC where the first primer is substantially complementary to the target
 CC sequence and the second primer is substantially complementary to the
 CC complement of the target sequence, and where the parts of the primers
 CC having substantial complementarity define the termini of the target
 CC sequence to be amplified, assay comprising contacting a test compound
 CC with the protein, and determining whether the test compound binds to the
 CC protein and a Streptococcus pneumoniae bacterium, where one or more
 CC genes encoding the proteins has been rendered inactive. The proteins,
 CC nucleic acid molecules, antibody and compositions are useful as
 CC medicaments for treating or preventing a disease or infection due to
 CC streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,
 CC sepsis, otitis media or ear infection. They are also useful in developing
 CC vaccines, diagnostics and antibiotics. The methods are useful for
 CC identifying immunodominant proteins. The present sequence is one of
 CC the 2469 identified coding region from the genomic sequence.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX Sequence 2457 BP; 836 A; 497 C; 531 G; 593 T; 0 other;

Query Match 2.2%; Score 53; DB 25; Length 2457;
 Best Local Similarity 100.0%; Pred. No. 3.9e-16;
 Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATGGGTACAGATTCAAG 1035
 DB 1000 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATGGGTACAGATTCAAG 1052

RESULT 16

ID AAA08557 standard; DNA; 2478 BP.
 AC AAA08557;
 DT 19-JUL-2000 (first entry)
 XX
 DE S. pneumoniae 92 kDa human C3-degrading protein coding sequence.
 XX
 KW Human C3-degrading protein; 92 kDa; immunostimulatory; vaccine;
 KM inhibitor; inflammation; organ rejection; xenotransplantation; ss.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200017370-A1.
 XX
 PD 30-MAR-2000.
 XX
 PF 24-SEP-1999; 99WO-US22362.
 XX
 PR 24-SEP-1998; 98US-0101736.
 PR 31-MAR-1999; 99US-0283094.
 XX
 PA (MNU) UNIV MINNESOTA.
 PA (ANCY) AMERICAN CYANAMID CO.
 XX
 PI Hostetter MK, Finkel DJ, Cheng Q, Green BA, Masi AM;
 XX
 DR WPI; 2000-283594/24.
 DR P-PSDB; AAY91939.
 XX
 PT Isolated polypeptide is used to stimulate immune system and immunize or
 PT treat a mammalian subject against Streptococcus pneumoniae infection or
 PT colonization
 XX
 PS Claim 1; Page 55-57; 63bp; English.
 XX
 CC The present sequence, isolated from Streptococcus pneumoniae, encodes a
 CC human C3-degrading protein (see AAY91939) of about 92 kDa. This
 CC sequence may encompass a smaller 20 kDa polypeptide coding sequence
 CC (AAA08556) also having human C3-degrading activity.
 CC The DNA sequences can be used for producing an immune response to
 CC Streptococcus pneumoniae in a mammal. Antibodies against the proteins
 CC can be used to inhibit S. pneumoniae-mediated C3 degradation.
 CC C3-mediated inflammation and rejection in xenotransplantation can be
 CC inhibited by expressing the nucleic acid sequences on the surface of an
 CC organ of an animal. In particular, the polypeptides are useful for
 CC stimulating the immune system and are effective to immunize or treat a
 CC mammalian subject against Streptococcus pneumoniae infection or
 CC colonization.
 XX
 SO Sequence 2478 BP; 837 A; 510 C; 535 G; 596 T; 0 other;

Query Match 2.2%; Score 53; DB 21; Length 2478;
 Best Local Similarity 100.0%; Pred. No. 3.9e-16;
 Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATGGGTACAGATTCAAG 1035
 DB 1003 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATGGGTACAGATTCAAG 1055

RESULT 17
AAA05417
ID AAA05417 standard; DNA; 2481 BP.
XX
AC AAA05417;
XX
DT 24-MAY-2000 (first entry)
XX
DE Streptococcus pneumoniae nucleotide sequence ID11.
XX
KM Streptococcus pneumoniae; vaccine; screening; protein antigen;
KM antibacterial; antiinflammatory; meningitis; infection; diagnosis;
KM pneumococcal disease; ds.
XX
OS Streptococcus pneumoniae.
XX
PN WO200006737-A2.
XX
PD 10-FEB-2000.
XX
PF 27-JUL-1999; 99WO-CB02451.
XX
PR 27-JUL-1998; 98GB-0016337.
PR 19-MAR-1999; 99US-0125164.
XX
PA (MICR-) MICROBIAL TECHNICS LTD.
XX
PI Gilbert CFG, Hansbro PM;
XX
DR WPI; 2000-1995300/17.
DR P-PSDB; AAY81662.
XX
PT New Streptococcal protein, useful as a vaccine, for diagnosis of
PT pneumococcal diseases and for screening agents capable of antagonizing
PT or inhibiting expression of the protein
XX
PS Claim 2; Page 99; 108pp; English.
XX
CC AAY81501 to AAY81679 represent specifically claimed protein sequences
CC isolated from Streptococcus pneumoniae. AAA05407 to AAA05590 represent
CC specifically claimed nucleotide sequences isolated from S. pneumoniae.
CC The sequences have antibacterial and antiinflammatory properties.
CC The protein sequences, and fragments of them, are useful as immunogens
CC and/or antigens. The nucleotide sequences can be used in vaccines and in
CC diagnostic assays. The proteins and nucleotides can be useful for the
CC detection and diagnosis of S. pneumoniae. The protein sequences are also
CC useful for screening an agent capable of antagonizing, inhibiting or
CC interfering with the function or expression of the proteins in which the
CC agent is useful for treatment or prophylaxis of S. pneumoniae infection
CC and meningitis. AAA05591 to AAA05614 represent primers used in the
CC exemplification of the present invention.
XX
SQ Sequence 2481 BP; 839 A; 514 C; 538 G; 590 T; 0 other;
Query Match 2.2%; Score 53; DB 21; Length 2481;
Best Local Similarity 100.0%; Pred. No. 3.9e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 983 GCTGTAATATTCCTTCCTGTTATGTTCAACCATTTGGGTACCGATTCAAG 1035
DB 1003 GCTGTAATATTCCTTCCTGTTATGTTCAACCATTTGGGTACCGATTCAAG 1055
RESULT 18
ABX06705
ID ABX06705 standard; DNA; 2517 BP.
XX
AC ABX06705;
XX
DT 11-FEB-2003 (first entry)
XX
DE S. pneumoniae type 4 strain coding region #993.
XX

KW Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media;
KW ear infection; antiinflammatory; antibacterial; immunostimulant;
KW auditory; respiratory; gene therapy; vaccine.
XX
OS Streptococcus pneumoniae type 4 strain.
XX
PN WO200277021-A2.
XX
PD 03-OCT-2002.
XX
PF 27-MAR-2002; 2002WO-IB02163.
XX
PR 27-MAR-2001; 2001GB-0007658.
XX
PA (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
PI Maignani V, Tettelin H, Fraser C;
XX
DR WPI; 2003-040579/03.
DR P-PSDB; ABU01418.
XX
PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT useful as medicaments for treating or preventing a disease or infection
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
PT or ear infection
XX
PS Claim 6; SEQ ID No 1985; 56pp; English.
XX
CC The invention relates to a protein comprising or having at least 50%
CC identity to any of the 2469 amino acid sequences, identified in the
CC specification (available on a computer readable format), or its fragment,
CC expressed from 2469 of 2489 identified DNA coding regions from the
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC ABS6454. Also included are an antibody which binds one of the
CC proteins, treating a patient by administering the protein, DNA or
CC antibody (in a composition), a kit comprising first and second primers,
CC which are the nucleic acid cited above or fragments between nucleotides
CC 8-100 of a sequence not defined in the specification, for amplifying a
CC target sequence contained within a Streptococcus nucleic acid sequence,
CC where the first primer is substantially complementary to the target
CC sequence and the second primer is substantially complementary to the
CC complement of the target sequence, and where the parts of the primers
CC having substantial complementarity define the termini of the target
CC sequence to be amplified, assay comprising contacting a test compound
CC with the protein, and determining whether the test compound binds to the
CC protein and a Streptococcus pneumoniae bacterium, where one or more
CC genes encoding the proteins has been rendered inactive. The proteins,
CC nucleic acid molecules, antibody and compositions are useful as
CC medicaments for treating or preventing a disease or infection due to
CC streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,
CC sepsis, otitis media or ear infection. They are also useful in developing
CC vaccines, diagnostics and antibiotics. The methods are useful for
CC identifying immunodominant proteins. The present sequence is one of
CC the 2489 identified coding region from the genomic sequence.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2517 BP; 830 A; 509 C; 545 G; 633 T; 0 other;
Query Match 2.2%; Score 53; DB 25; Length 2517;
Best Local Similarity 100.0%; Pred. No. 4e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 983 GCTGTAATATTCCTTCCTGTTATGTTCAACCATTTGGGTACCGATTCAAG 1035
DB 1003 GCTGTAATATTCCTTCCTGTTATGTTCAACCATTTGGGTACCGATTCAAG 1055
RESULT 19
AAA47602

```
ID AAA47602 standard; DNA; 2531 BP.
XX
AC AAA47602;
XX
DT 20-OCT-2000 (first entry)
XX
DE Recombinant variant of Sp36 gene (Sp36D) of S. pneumoniae.
XX
KW Streptococcus pneumoniae; infection; vaccine; coiled coil region;
KW histidine triad residue; Sp36; antibody; otitis media;
KW nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
KW meningitis; lobar pneumonia; ds.
XX
CS Streptococcus pneumoniae.
XX
FH Key Location/Qualifiers
FT CDS 1..2517
FT /tag= a
FT /product= Sp36D polypeptide
XX
PN WO200037105-A2.
XX
PD 29-JUN-2000.
XX
PF 21-DEC-1999; 99WO-US30390.
XX
PR 21-DEC-1998; 98US-0113048.
XX
PA (MEDI-) MEDIMMUNE INC.
XX
PI Johnson LS, Koenig S, Adamou JB;
XX
DR WPI: 2000-452129/39.
DR P-PSDB; AAB01466.
XX
PT Vaccine useful for prophylaxis and treatment of pneumococcal infections
PT such as otitis media, nasopharyngeal and bronchial infections,
PT comprises Streptococcus pneumoniae proteins
XX
PS Disclosure; Page 57-58; 70pp; English.
XX
CC Although a number of proteins have been suggested as being involved
CC in the pathogenicity of Streptococcus pneumoniae, there still remains
CC a need to identify polypeptides having epitopes in common from
CC various strains of S. pneumoniae in order to utilise such
CC polypeptides in vaccines to protect against a wide variety of
CC S. pneumoniae. New vaccine compositions are described which comprise a
CC Streptococcus pneumoniae polypeptide (or fragments) of 80 - 680 amino
CC acids in length that comprise at least one histidine triad residue
CC (HxxHxH) or a coiled-coil region, or an antibody directed against
CC these features. The vaccine is useful in protecting against infection
CC by Streptococcus pneumoniae. The vaccine composition comprising
CC antibodies to is useful for passive immunization for treating
CC pneumococcal infections which includes otitis media, nasopharyngeal
CC and bronchial infections.
XX
SQ Sequence 2531 BP; 836 A; 511 C; 547 G; 637 T; 0 other;
XX
Query Match 2.2%; Score 53; DB 21; Length 2531;
Best Local Similarity 100.0%; Pred. No. 4e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 983 GCTGTAATATTCCTTCGTTATCGTTCAACCACTGGGATTCAGATTCAAG 1035
DB 1003 GCTGTAATATTCCTTCGTTATCGTTCAACCACTGGGATTCAGATTCAAG 1055
XX
RESULT 20
AAA47605
ID AAA47605 standard; DNA; 2531 BP.
XX
AC AAA47605;
XX
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae BWH-11-2 gene SEQ ID NO:13.
XX
```

```
DT 20-OCT-2000 (first entry)
XX
DE Recombinant variant of Sp36 gene (Sp36B) of S. pneumoniae.
XX
KW Streptococcus pneumoniae; infection; vaccine; coiled coil region;
KW histidine triad residue; Sp36; antibody; otitis media;
KW nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
KW meningitis; lobar pneumonia; ds.
XX
CS Streptococcus pneumoniae.
XX
FH Key Location/Qualifiers
FT CDS 1..2531
FT /tag= a
FT /product= Sp36B polypeptide
XX
PN WO200037105-A2.
XX
PD 29-JUN-2000.
XX
PF 21-DEC-1999; 99WO-US30390.
XX
PR 21-DEC-1998; 98US-0113048.
XX
PA (MEDI-) MEDIMMUNE INC.
XX
PI Johnson LS, Koenig S, Adamou JB;
XX
DR WPI: 2000-452129/39.
DR P-PSDB; AAB01469.
XX
PT Vaccine useful for prophylaxis and treatment of pneumococcal infections
PT such as otitis media, nasopharyngeal and bronchial infections,
PT comprises Streptococcus pneumoniae proteins
XX
PS Disclosure; Page 69-70; 70pp; English.
XX
CC Although a number of proteins have been suggested as being involved
CC in the pathogenicity of Streptococcus pneumoniae, there still remains
CC a need to identify polypeptides having epitopes in common from
CC various strains of S. pneumoniae in order to utilise such
CC polypeptides in vaccines to protect against a wide variety of
CC S. pneumoniae. New vaccine compositions are described which comprise a
CC Streptococcus pneumoniae polypeptide (or fragments) of 80 - 680 amino
CC acids in length that comprise at least one histidine triad residue
CC (HxxHxH) or a coiled-coil region, or an antibody directed against
CC these features. The vaccine is useful in protecting against infection
CC by Streptococcus pneumoniae. The vaccine composition comprising
CC antibodies to is useful for passive immunization for treating
CC pneumococcal infections which includes otitis media, nasopharyngeal
CC and bronchial infections.
XX
SQ Sequence 2531 BP; 861 A; 508 C; 550 G; 609 T; 3 other;
XX
Query Match 2.2%; Score 53; DB 21; Length 2531;
Best Local Similarity 100.0%; Pred. No. 4e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 983 GCTGTAATATTCCTTCGTTATCGTTCAACCACTGGGATTCAGATTCAAG 1035
DB 1000 GCTGTAATATTCCTTCGTTATCGTTCAACCACTGGGATTCAGATTCAAG 1052
XX
RESULT 21
AAA65737
ID AAA65737 standard; DNA; 2639 BP.
XX
AC AAA65737;
XX
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae BWH-11-2 gene SEQ ID NO:13.
XX
```

KM Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KM prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KM otitis media; pneumonia; immunisation; bactericidal; ds.
 OS Streptococcus pneumoniae.
 XX
 XX WO200039299-A2.
 XX
 XX
 XX PD 06-JUL-2000.
 XX
 XX PF 20-DEC-1999; 99WO-CA01218.
 XX
 XX PR 23-DEC-1998; 98US-0113800.
 XX
 XX PA (BIOC-) BIOCHEM PHARMA INC.
 XX
 XX PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 DR WPI; 2000-452397/39.
 DR P-PSDB; AAB12720.
 XX
 XX PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteraemia and/or pneumonia -
 PS Example 6; Fig 16; 106pp; English.
 XX
 CC The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the protein
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence encodes the
 CC S. pneumoniae BVH-11-2 protein antigen.
 XX
 XX SQ Sequence 2639 BP; 889 A; 518 C; 567 G; 665 T; 0 other;
 Query Match 2.2%; Score 53; DB 21; Length 2639;
 Best Local Similarity 100.0%; Pred. No. 4e-16;
 Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 983 GCTCGTATTATTCCTTCGTTATGTTCAACCATTTGGTACGATTCAAG 1035
 DB 1113 GCTCGTATTATTCCTTCGTTATGTTCAACCATTTGGTACGATTCAAG 1165
 RESULT 22
 ABK15104
 ID ABK15104 standard; DNA; 2639 BP.
 XX
 AC ABK15104;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 XX DNA encoding Streptococcus pneumoniae BVH-11-2.
 DE
 XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 KM pneumonia; streptococcal bacterial infection; gene; ds; BVH-11-2.
 XX
 OS Streptococcus pneumoniae.
 XX
 XX Key Location/Qualifiers
 FH CDS 114..2630
 FT /*tag= a
 FT /product= "BVH-11-2"
 FT /note= "The gene is flanked by sequences from the
 FT vector SP64, no information on which is
 FT given in the specification"
 XX
 XX WO200198334-A2.
 XX
 XX PD 27-DEC-2001.

XX
 XX 19-JUN-2001; 2001WO-CA00908.
 PF
 XX
 XX 20-JUN-2000; 2000US-212683P.
 PR
 XX
 XX (SHIR-) SHIRE BIOCHEM INC.
 PA
 XX
 XX Hamel J, Queller C, Charland N, Martin D, Brodeur B;
 PI
 XX
 XX WPI; 2002-122272/16.
 DR
 XX P-PSDB; AA075934.
 DR
 XX
 XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia -
 PS Example 3; Fig 5; 113pp; English.
 XX
 XX
 CC The invention describes an isolated polypeptide (I) with 70-90%
 CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novacardia or
 CC Streptococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence encodes the
 CC Streptococcus pneumoniae protein BVH-11-2, used to create the antigenic
 CC peptides described in the method of the invention.
 XX
 XX SQ Sequence 2639 BP; 889 A; 518 C; 567 G; 665 T; 0 other;
 Query Match 2.2%; Score 53; DB 24; Length 2639;
 Best Local Similarity 100.0%; Pred. No. 4e-16;
 Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 983 GCTCGTATTATTCCTTCGTTATGTTCAACCATTTGGTACGATTCAAG 1035
 DB 1113 GCTCGTATTATTCCTTCGTTATGTTCAACCATTTGGTACGATTCAAG 1165
 RESULT 23
 ABS56454
 ID ABS56454 standard; DNA; 2162598 BP.
 XX
 AC ABS56454;
 XX
 DT 10-FEB-2003 (first entry)
 XX
 XX Streptococcus pneumoniae type 4 strain complete genome.
 DE
 XX ds; bacterial meningitis; pneumonia; sepsis; otitis media; genome;
 KM ear infection; antiinflammatory; antibacterial; immunostimulant;
 KM auditory; respiratory; gene therapy; vaccine.
 XX
 OS Streptococcus pneumoniae type 4 strain.
 XX
 XX WO200277021-A2.
 PN
 XX 03-OCT-2002.
 XX
 XX PD 27-MAR-2002; 2002WO-IB02163.
 XX
 XX

PR 27-MAR-2001; 2001GB-0007658.
XX
XX (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
XX Maignani V, Tettein H, Fraser C;
XX WPI; 2003-040579/03.
DR
XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT useful as medicaments for treating or preventing a disease or infection
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
PT or ear infection
XX
XX Claim 17; SEQ ID No 4979; 56pp; English.
XX
XX The invention relates to a protein comprising or having at least 50%
CC identity to any of the 2469 amino acid sequences, identified in the
CC specification (available on a computer readable format), or its fragment,
CC expressed from 2469 of 2489 identified DNA coding regions from the
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC AB556454. Also included are an antibody which binds one of the
CC proteins, treating a patient by administering the protein, DNA or
CC antibody (in a composition), a kit comprising first and second primers,
CC which are the nucleic acid cited above or fragments between nucleotides
CC 8-100 of a sequence not defined in the specification, for amplifying a
CC target sequence contained within a Streptococcus nucleic acid sequence,
CC where the first primer is substantially complementary to the target
CC sequence and the second primer is substantially complementary to the
CC complement of the target sequence, and where the parts of the primers
CC having substantial complementarity define the termini of the target
CC sequence to be amplified, assay comprising contacting a test compound
CC with the protein, and determining whether the test compound binds to the
CC protein and a Streptococcus pneumoniae bacterium, where one or more
CC genes encoding the proteins has been rendered inactive. The proteins,
CC nucleic acid molecules, antibody and compositions are useful as
CC medicaments for treating or preventing a disease or infection due to
CC streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,
CC sepsis, otitis media or ear infection. They are also useful in developing
CC vaccines, diagnostics and antibiotics. The methods are useful for
CC identifying immunodominant proteins. The present sequence is the
CC Streptococcus pneumoniae type 4 strain genome sequence.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pcf_sequences.
XX
XX Sequence 2162598 BP; 654373 A; 427176 C; 431369 G; 649680 T; 0 other;
SQ
XX
XX Query Match 2.1%; Score 53; DB 25; Length 2162598;
Best Local Similarity 100.0%; Pred. No. 5.1e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 983 GCTGATATTATCCCTTCGTTATCGTTCAACCATTTGGGTACGAGATTCAAG 1035
Db 838951 GCTGATATTATCCCTTCGTTATCGTTCAACCATTTGGGTACGAGATTCAAG 839003
RESULT 24
AAV52393
ID AAV52393 standard; DNA; 492 BP.
XX
XX AAV52393;
AC
XX 19-JUL-1999 (first entry)
DT
XX
XX Streptococcus pneumoniae complement C3-degrading protease DNA.
DE
XX
XX Human complement C3-degrading protease; vaccine; infection;
KM meningitis; pneumonia; xerotrapiplantation; transplant rejection;
KM inflammation; ds.
XX
XX Streptococcus pneumoniae.
OS

XX
XX W09915675-A1.
XX
XX 01-APR-1999.
ED
XX
XX 24-SEP-1998; 98WO-US20186.
XX
XX 24-SEP-1997; 97US-0059907.
PR
XX
XX (AMCY) AMERICAN CYANAMID CO.
PA (MINN) UNITV MINNESOTA.
XX
XX Cheng Q, Finkel DJ, Green BA, Hostetter MK, Masi AW;
XX WPI; 1999-254719/21.
XX P-PSDB; AAV05752.
DR
XX
XX New isolated human complement C3-degrading proteinase
PT
XX
XX Claim 21; Page 49; 66pp; English.
XX
XX This DNA sequence encodes a 20 kDa protein (see AAV05752) of
CC Streptococcus pneumoniae serotype 3 that is capable of degrading
CC human complement protein C3 (HCP3). It was isolated from a
CC library of insertionally interrupted pneumococcal genes by
CC identifying those clones that had decreased C3 degrading activity
CC as compared to wild-type S. pneumoniae. Alignment of the sequence
CC with the S. pneumoniae (serotype 4) genome showed that the open
CC reading frame that codes for the 20 kDa protein may be part of a
CC larger open reading frame, i.e. a 2163 bp (see AAV52394) in the
CC serotype 4 genome, which encodes a protein of approximately 79
CC kDa (see AAV05753). Amino acids 1-58 and 90-132 of the 20 kDa
CC protein have substantial sequence identity with amino acids 170-227
CC and 258-300 of the 79 kDa protein. Proteins and peptides or
CC polypeptides containing these regions, and DNA sequences encoding
CC them (nucleotides 1-174 and 320-492 of the present sequence) are
CC claimed. HCP3 proteases and polypeptides can be used as immune
CC system stimulating compositions (claimed). They can produce an
CC immune response against S. pneumoniae to immunize or treat a
CC mammalian subject against infection or colonization (claimed).
CC They can produce a B cell response, a T cell response, an
CC epithelial cell response, or an endothelial cell response
CC (claimed). The expression of the proteins on the surface of an
CC organ of an animal used in xenotransplantation can be used to
CC inhibit C3-mediated inflammation and rejection.
XX
XX Sequence 492 BP; 152 A; 119 C; 98 G; 123 T; 0 other;
SQ
XX
XX Query Match 2.1%; Score 50; DB 20; Length 492;
Best Local Similarity 100.0%; Pred. No. 1.2e-14;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 985 TCGATATTATCCCTTCGTTATCGTTCAACCATTTGGGTACGAGATTCA 1034
Db 185 TCGATATTATCCCTTCGTTATCGTTCAACCATTTGGGTACGAGATTCA 234
RESULT 25
AAV52488
ID AAV52488 standard; DNA; 973 BP.
XX
XX AAV52488;
AC
XX 23-OCT-1998 (first entry)
DT
XX
XX Streptococcus pneumoniae genome fragment SEQ ID NO:355.
DE
XX
XX Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KM computer readable medium; vaccine; pharmaceutical composition; ds.
XX
XX Streptococcus pneumoniae.
XX
XX W09818931-A2.
PN

XX The present sequence represents a cloning primer used in an example from
 CC the present invention which describes proteins from Streptococcus
 CC pneumoniae. Nucleic acid sequence encoding Streptococcus pneumoniae
 CC proteins can be useful in vaccines for inducing protective antibodies
 CC against Streptococcus pneumoniae, for treatment or prevention of
 CC infection e.g. pneumonia, otitis media or meningitis. Probes based on
 CC the nucleic acids are used to detect Streptococcus infection (by usual
 CC hybridisation or amplification methods), also for isolating
 CC Streptococcus genes or their allelic variants. The proteins can be used
 CC similarly to detect specific antibodies in standard immunoassays,
 CC especially for diagnosing or monitoring infections. Antibodies which
 CC bind the proteins are used to detect corresponding antigens, to purify
 CC the proteins and for passive immunisation (optionally coupled to a
 CC toxin). Vaccines are administered, e.g. by injection, orally or through
 CC the skin, typically at 0.01-1000 (especially 10-300) mu g/ml per dose.
 CC The cloning primers used in the present invention are given in AAV27437
 CC to AAV27562 and AAV39870 to AAV39969.
 CC XX
 SQ Sequence 40 BP; 10 A; 6 C; 7 G; 17 T; 0 other;
 Query Match 1.3%; Score 30; DB 19; Length 40;
 Best Local Similarity 100.0%; Pred. No. 0.00016;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 2360 CCTCATCTGTAAGTAAAGAAAAATAAAC 2389
 DB 40 CCTCATCTGTAAGTAAAGAAAAATAAAC 11
 RESULT 28
 ID AAA47601/c
 AC AAA47601 standard; DNA; 40 BP.
 XX
 XX AAA47601;
 DT 20-OCT-2000 (first entry)
 XX
 DE Primer for amplifying Sp36 gene of S. pneumoniae.
 XX
 KW Streptococcus pneumoniae; infection; vaccine; coiled coil region;
 KW histidine triad residue; Sp36; antibody; otitis media;
 KW nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
 KW meningitis; lobar pneumonia; primer; ss.
 XX
 OS Synthetic.
 XX
 PN WO200037105-A2.
 XX
 PD 29-JUN-2000.
 XX
 PF 21-DEC-1999; 99WO-US30390.
 XX
 PR 21-DEC-1998; 98US-0113048.
 XX
 PA (MEDI-) MEDIMMUNE INC.
 XX
 PI Johnson US, Koenig S, Adamou JE;
 XX
 DR WPI: 2000-452129/39.
 PT Vaccine useful for prophylaxis and treatment of pneumococcal infections
 PT such as otitis media, nasopharyngeal and bronchial infections,
 PT comprises Streptococcus pneumoniae proteins
 XX
 PS Example 1; Page 54; 70pp; English.
 XX
 CC Although a number of proteins have been suggested as being involved
 CC in the pathogenicity of Streptococcus pneumoniae, there still remains
 CC a need to identify polypeptides having epitopes in common from
 CC various strains of S. pneumoniae in order to utilise such
 CC polypeptides in vaccines to protect against a wide variety of
 CC S. pneumoniae. New vaccine compositions are described which comprise a

CC Streptococcus pneumoniae polypeptide (or fragments) of 80 - 680 amino
 CC acids in length that comprise at least one histidine triad residue
 CC (HxxHxH) or a coiled-coil region, or an antibody directed against
 CC these features. The vaccine is useful in protecting against infection
 CC by Streptococcus pneumoniae. The vaccine composition comprising
 CC antibodies to is useful for passive immunization for treating
 CC pneumococcal infections which includes otitis media, nasopharyngeal
 CC and bronchial infections. Three primers (AAA47599-601) were designed
 CC that would amplify the Sp36 gene and allow its cloning into pGSI10
 CC and expression as a histidine tagged protein lacking the signal
 CC sequence for purification by nickel-affinity chromatography.
 CC XX
 SQ Sequence 40 BP; 10 A; 6 C; 7 G; 17 T; 0 other;
 Query Match 1.3%; Score 30; DB 21; Length 40;
 Best Local Similarity 100.0%; Pred. No. 0.00016;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 2360 CCTCATCTGTAAGTAAAGAAAAATAAAC 2389
 DB 40 CCTCATCTGTAAGTAAAGAAAAATAAAC 11
 RESULT 29
 ID ABQ84960/c
 AC ABQ84960 standard; DNA; 40 BP.
 XX
 XX ABQ84960;
 DT 04-SEP-2002 (first entry)
 XX
 DE Streptococcus pneumoniae ORF cloning primer SEQ ID NO:282.
 XX
 KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
 KW antibacterial; Streptococcal infection; detection; primer; ss.
 XX
 OS Streptococcus pneumoniae.
 OS Synthetic.
 XX
 PN US2002061545-A1.
 XX
 PD 23-MAY-2002.
 XX
 PF 22-JAN-2001; 2001US-0765272.
 XX
 PR 30-OCT-1997; 97US-0961083.
 XX
 PA (CHOI/) CHOI G H.
 PA (KUNS/) KUNSCH C A.
 PA (BARA/) BARASH S C.
 PA (DILL/) DILLON P J.
 PA (DOUG/) DOUGHERTY B.
 PA (FANN/) FANNON M R.
 PA (ROSE/) ROSEN C A.
 XX
 PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
 PI Rosen CA;
 XX
 DR WPI: 2002-479261/51.
 PT New Streptococcus pneumoniae antigens, useful for detecting
 PT Streptococcus and for preventing or attenuating disease caused by
 PT Streptococcus infection -
 XX
 PS Example 1; Page 63; 70pp; English.
 XX
 CC ABQ84792 to ABQ84904 represents nucleic acids which encode the
 CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669.
 CC The S. pneumoniae antigens have antibacterial activity and can be
 CC used in vaccines. The S. pneumoniae antigens can also be used to
 CC prevent or attenuate a Streptococcal infection in an animal. The
 CC polynucleotides encoding the S. pneumoniae antigens can be used to
 CC detect Streptococcus nucleic acids. ABQ84905 to ABQ85130 represent

CC primers used in the cloning of *S. pneumoniae* ORFs (open reading frames)
 CC which are used in an example from the present invention.
 XX
 SQ Sequence 40 BP; 10 A; 6 C; 7 G; 17 T; 0 other;

Query Match 1.3%; Score 30; DB 24; Length 40;
 Best Local Similarity 100.0%; Pred. No. 0.00016;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2360 CCTCATCTGTAGTAAAGGAAAAATTAAC 2389
 DB 40 CCTCATCTGTAGTAAAGGAAAAATTAAC 11

RESULT 30
 AAA47599
 ID AAA47599 standard; DNA; 36 BP.

AC AAA47599;

DT 20-OCT-2000 (first entry)

DE Primer for amplifying Sp36 gene of *S. pneumoniae*.

XX Streptococcus pneumoniae; infection; vaccine; coiled coil region;
 KW histidine triad residue; Sp36; antibody; otitis media;
 KW nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
 KW meningitis; lobar pneumonia; primer; ss.

OS Synthetic.

PN WO200037105-A2.

PD 29-JUN-2000.

PF 21-DEC-1999; 99WO-US30390.

PR 21-DEC-1998; 98US-0113048.

PA (MEDI-) MEDIMUNE INC.

PI Johnson LS, Koenig S, Adamou JE;

DR WPI; 2000-452129/39.

XX
 PT Vaccine useful for prophylaxis and treatment of pneumococcal infections
 PT such as otitis media, nasopharyngeal and bronchial infections,
 PT comprises Streptococcus pneumoniae proteins
 XX
 PS Example 1; Page 53; 70pp; English.

XX Although a number of proteins have been suggested as being involved
 CC in the pathogenicity of Streptococcus pneumoniae, there still remains
 CC a need to identify polypeptides having epitopes in common from
 CC various strains of *S. pneumoniae* in order to utilize such
 CC polypeptides in vaccines to protect against a wide variety of
 CC *S. pneumoniae*. New vaccine compositions are described which comprise a
 CC Streptococcus pneumoniae polypeptide (or fragments) of 80 - 680 amino
 CC acids in length that comprise at least one histidine triad residue
 CC (HxxHxx) or a coiled-coil region, or an antibody directed against
 CC these features. The vaccine is useful in protecting against infection
 CC by Streptococcus pneumoniae. The vaccine composition comprising
 CC antibodies to its useful for passive immunization for treating
 CC pneumococcal infections which includes otitis media, nasopharyngeal
 CC and bronchial infections. Three primers (AAA47599-601) were designed
 CC that would amplify the Sp36 gene and allow its cloning into pOE10
 CC and expression as a histidine tagged protein lacking the signal
 CC sequence for purification by nickel-affinity chromatography.
 XX

Sequence 36 BP; 8 A; 8 C; 9 G; 11 T; 0 other;

Query Match 1.1%; Score 27; DB 21; Length 36;
 Best Local Similarity 100.0%; Pred. No. 0.0054;

Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTTACGAGTTGGAGCTGTATCAAGC 27
 DB 10 TTCTTACGAGTTGGAGCTGTATCAAGC 36

RESULT 31
 AAV27491
 ID AAV27491 standard; DNA; 37 BP.

AC AAV27491;

DT 02-OCT-1998 (first entry)

DE Streptococcus pneumoniae ORF cloning primer SEQ ID NO:281.

XX Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
 KW detection; pneumonia; otitis media; meningitis; cloning primer; ss.

OS Synthetic.

PN Streptococcus pneumoniae.

PN WO9818930-A2.

PD 07-MAY-1998.

PF 30-OCT-1997; 97WO-US19422.

PR 31-OCT-1996; 96US-0029960.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Choi GH, Hromocky J A, Johnson LS, Kunsch CA;

DR WPI; 1998-272224/24.

XX
 PT Nucleic acid encoding antigenic peptide(s) from Streptococcus
 PT pneumoniae - or their epitope-containing fragments, useful in
 PT protective or therapeutic vaccines, and for diagnosis
 XX
 PS Example 1; Page 106; 118pp; English.

XX The present sequence represents a cloning primer used in an example from
 CC the present invention which describes proteins from Streptococcus
 CC pneumoniae. Nucleic acid sequence encoding Streptococcus pneumoniae
 CC proteins can be useful in vaccines for inducing protective antibodies
 CC against Streptococcus pneumoniae, for treatment or prevention of
 CC infection e.g. pneumonia, otitis media or meningitis. Probes based on
 CC the nucleic acids are used to detect Streptococcus infection (by usual
 CC hybridisation or amplification methods), also for isolating
 CC Streptococcus genes or their allelic variants. The proteins can be used
 CC especially for detecting specific antibodies in standard immunoassays,
 CC bind the proteins are used to detect corresponding antigens, to purify
 CC the proteins and for passive immunisation (optionally coupled to a
 CC toxin). Vaccines are administered, e.g. by injection, orally or through
 CC the skin, typically at 0.01-1000 (especially 10-300) mu g/ml per dose.
 CC The cloning primers used in the present invention are given in AAV27437
 CC to AAV27562 and AAV29870 to AAV39969.
 XX

Sequence 37 BP; 8 A; 8 C; 10 G; 11 T; 0 other;

Query Match 1.1%; Score 27; DB 19; Length 37;
 Best Local Similarity 100.0%; Pred. No. 0.0054;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTTACGAGTTGGAGCTGTATCAAGC 27
 DB 11 TTCTTACGAGTTGGAGCTGTATCAAGC 37

RESULT 32

ABQ84959	ID	ABQ84959 standard; DNA; 37 BP.
XX AC	ABQ84959;	
XX XX	04-SEP-2002 (first entry)	
XX DT	Streptococcus pneumoniae ORF cloning primer SEQ ID NO:281.	
XX DE	Streptococcus pneumoniae; epitope; vaccine; antigenic protein;	
XX KW	antibacterial; Streptococcal infection; detection; primer; ss.	
XX OS	Streptococcus pneumoniae.	
XX OS	Synthetic.	
XX PN	US2002061545-A1.	
XX PD	23-MAY-2002.	
XX PF	22-JAN-2001; 2001US-0765272.	
XX PR	30-OCT-1997; 97US-0961083.	
XX PA	(CHOI/) CHOI G H.	
XX PA	(KUNS/) KUNSCH C A.	
XX PA	(BARA/) BARASH S C.	
XX PA	(DILL/) DILLON P J.	
XX PA	(DOUG/) DOUGHERTY B.	
XX PA	(FANN/) FANNON M R.	
XX PA	(ROSE/) ROSEN C A.	
XX PI	Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR,	
XX PI	Rosen CA;	
XX DR	WP1; 2002-479261/51.	
XX PT	New Streptococcus pneumoniae antigens, useful for detecting	
XX PT	Streptococcus and for preventing or attenuating disease caused by	
XX PS	Streptococcus infection -	
XX PS	Example 1; Page 63; 70pp; English.	
XX CC	ABQ84792 to ABQ84904 represents nucleic acids which encode the	
XX CC	Streptococcus pneumoniae antigens given in ABP54557 to ABP54669.	
XX CC	The S. pneumoniae antigens have antibacterial activity and can be	
XX CC	used in vaccines. The S. pneumoniae antigens can also be used to	
XX CC	prevent or attenuate a Streptococcal infection in an animal. The	
XX CC	polynucleotides encoding the S. pneumoniae antigens can be used to	
XX CC	detect Streptococcus nucleic acids. ABQ84905 to ABQ85130 represent	
XX CC	primers used in the cloning of S. pneumoniae ORFs (open reading frames)	
XX CC	which are used in an example from the present invention.	
XX SQ	Sequence 37 BP; 8 A; 8 C; 10 G; 11 T; 0 other;	
OY	Query Match 1.1%; Score 27; DB 24; Length 37;	
	Best Local Similarity 100.0%; Pred. No. 0.0054;	
	Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Db	1 TTCTTAGCAGTTGGACTGTATCAAGC 27	
	11 TTCTTACGAGTTGGACTGTATCAAGC 37	
RESULT 33		
AAA47600	ID AAA47600 standard; DNA; 35 BP.	
XX AC	AAA47600;	
XX DT	20-OCT-2000 (first entry)	
XX XX	Primer for amplifying Sp36 gene of S. pneumoniae.	
XX XX		

KW Streptococcus pneumoniae; infection; vaccine; coiled coil region;
KW histidine triad residue; Sp36 antibody; otitis media;
KM nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
KW meningitis; lobar pneumonia; primer; ss.
XX
XX Synthetic.
OS
PN W0200037105-A2.
PD
XX 29-JUN-2000.
XX
XX 21-DEC-1999; 99WO-US30390.
PF
XX 21-DEC-1998; 98US-0113048.
PR
PA (MED1-) MEDIMUNE INC.
PI Johnson LS, Koenig S, Adamou JE;
XX WPI: 2000-452129/39.
DR
XX Vaccine useful for prophylaxis and treatment of pneumococcal infections
PT such as otitis media, nasopharyngeal and bronchial infections,
PT comprises Streptococcus pneumoniae proteins
XX
XX Example 1; Page 53; 70pp; English.
PS
CC Although a number of proteins have been suggested as being involved
CC in the pathogenicity of Streptococcus pneumoniae, there still remains
CC a need to identify polypeptides having epitopes in common from
CC various strains of S. pneumoniae in order to utilise such
CC polypeptides in vaccines to protect against a wide variety of
CC S. pneumoniae. New vaccine compositions are described which comprise a
CC Streptococcus pneumoniae polypeptide (or fragments) of 80 - 680 amino
CC acids in length that comprise at least one histidine triad residue
CC (HxxHxN) or a coiled-coil region, or an antibody directed against
CC these features. The vaccine is useful in protecting against infection
CC by Streptococcus pneumoniae. The vaccine composition comprising
CC antibodies to is useful for passive immunization for treating
CC Pneumococcal infections which includes otitis media, nasopharyngeal
CC and bronchial infections. Three primers (AAA47599-601) were designed
CC that would amplify the Sp36 gene and allow its cloning into pCR10
CC and expression as a histidine tagged protein lacking the signal
CC sequence for purification by nickel-affinity chromatography.
CC
XX
XX Sequence 35 BP; 11 A; 7 C; 9 G; 8 T; 0 other;
SQ

Query March 1.1k; Score 26; DB 21; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.017;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 16 ACTGTATCAAGCTAGAACGGTTAAGG 41
|||
DB 10 ACTGTATCAAGCTAGAACGGTTAAGG 35
|||||

RESULT 34
ID AAA65762
AAA65762 standard; DNA; 33 BP.

AA65762;
XX
XX 21-NOV-2000 (first entry)
DE Streptococcus pneumoniae BVH-11 PCR primer SEQ ID NO:37.
XX
XX Streptococcus pneumoniae, BVH-3; BVH-11; BVH-28; antigen; vaccine;
KM prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal; PCR primer; ss.
XX
XX Streptococcus pneumoniae.
OS
XX
PN W0200039299-A2.

```
XX 06-JUL-2000.
PD 20-DEC-1999; 99WO-CA01218.
XX 23-DEC-1998; 98US-0113800.
XX (BIOC-) BIOCHEM PHARMA INC.
XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI; 2000-452397/39.
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX otitis media, bacteraemia and/or pneumonia -
XX Example 8; Page 49; 106pp; English.
XX The present invention describes nucleic acids (I) encoding protein
XX antigens (II) from Streptococcus pneumoniae. The protein antigens
XX have bactericidal activity. The nucleic acids, encoding the protein
XX antigens, may be used for the recombinant production of the proteins
XX they encode. The protein antigens may then be used as vaccines for the
XX prevention and treatment of Streptococcal infections in mammals
XX (especially humans) which result in, e.g. meningitis, otitis media,
XX bacteraemia and/or pneumonia. The present sequence represents a PCR
XX primer for a S. pneumoniae GAS BVH-71 protein antigen, which is used in
XX an example from the present invention.
SQ Sequence 33 BP; 7 A; 9 C; 7 G; 10 T; 0 other;
QY Query Match 1.0%; Score 25; DB 21; Length 33;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 998 CTTGTTATCGTTCAACCATTTGGG 1022
CTTCGTTATCGTTCAACCATTTGGG 33
QY 998 CTTGTTATCGTTCAACCATTTGGG 1022
CTTCGTTATCGTTCAACCATTTGGG 33
DB 9 CTTGTTATCGTTCAACCATTTGGG 33
RESULT 35
ABK33684
ID ABK33684 standard; DNA; 33 BP.
XX
XX ABK33684;
AC 08-MAY-2002 (first entry)
XX
XX S. pneumoniae BVH-11 gene, PCR primer HAMJ 282.
XX
XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
XX pneumonia; Streptococcal bacterial infection; PCR; primer; ss.
XX
XX Streptococcus pneumoniae.
OS
XX WO200198334-A2.
XX
XX 27-DEC-2001.
PD
XX
XX 19-JUN-2001; 2001WO-CA00908.
PF
XX
XX 20-JUN-2000; 2000US-212683P.
PR
XX
XX (SHIR-) SHIRE BIOCHEM INC.
PA
XX
XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX WPI; 2002-122272/16.
XX
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX epitope-bearing polypeptides, useful as vaccine components for treating
XX or preventing streptococcal infections such as otitis media,
XX meningitis, and bacteraemia
```

```
XX Example 1; Page 31; 113pp; English.
XX
XX The invention describes an isolated polypeptide (I) with 70-90%
XX identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
XX BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
XX comprising (I) is useful for therapeutic or prophylactic treatment of
XX meningitis, otitis media, bacteraemia or pneumonia infection in an
XX individual susceptible to these disorders. (II) is also useful for
XX therapeutic or prophylactic treatment of any streptococcal bacterial
XX infection (e.g., caused by Streptococcus pneumoniae, group A
XX Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
XX as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novae or
XX Staphylococcus aureus) in an individual susceptible to the infection.
XX A polynucleotide (III) encoding (I) is useful in DNA immunisation
XX techniques. The Streptococcus polypeptides are useful in a diagnostic
XX test for S. pneumoniae infection. (III) is useful for designing DNA
XX probes for use in detecting the presence of Streptococcus in a biological
XX sample suspected of containing the bacteria. The DNA probes may also be
XX used for detecting circulating S. pneumonia nucleic acid in a sample for
XX diagnosing streptococcal infections. This sequence represents a primer
XX used for the isolation of S. pneumoniae genes from which the antigenic
XX peptides of the invention are derived.
SQ Sequence 33 BP; 7 A; 9 C; 7 G; 10 T; 0 other;
QY Query Match 1.0%; Score 25; DB 24; Length 33;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 998 CTTGTTATCGTTCAACCATTTGGG 1022
CTTCGTTATCGTTCAACCATTTGGG 33
QY 998 CTTGTTATCGTTCAACCATTTGGG 1022
CTTCGTTATCGTTCAACCATTTGGG 33
DB 9 CTTGTTATCGTTCAACCATTTGGG 33
RESULT 36
AAA65764
ID AAA65764 standard; DNA; 34 BP.
XX
XX AAA65764;
AC 21-NOV-2000 (first entry)
XX
XX Streptococcus pneumoniae BVH-11 PCR primer SEQ ID NO:39.
XX
XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
XX prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX otitis media; pneumonia; immunisation; bactericidal; PCR primer; ss.
XX
XX Streptococcus pneumoniae.
OS
XX WO2000039299-A2.
XX
XX 06-JUL-2000.
PD
XX
XX 20-DEC-1999; 99WO-CA01218.
PF
XX
XX 23-DEC-1998; 98US-0113800.
PR
XX
XX (BIOC-) BIOCHEM PHARMA INC.
PA
XX
XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI; 2000-452397/39.
XX
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX otitis media, bacteraemia and/or pneumonia -
XX Example 8; Page 49; 106pp; English.
XX The present invention describes nucleic acids (I) encoding protein
XX antigens (II) from Streptococcus pneumoniae. The protein antigens
XX have bactericidal activity. The nucleic acids, encoding the protein
```

CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents a PCR
CC primer for a S. pneumoniae GAS BVH-71 protein antigen, which is used in
CC an example from the present invention.
XX
SQ Sequence 34 BP; 10 A; 8 C; 7 G; 9 T; 0 other;
Query Match 1.0%; Score 25; DB 21; Length 34;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 797 GACATTGATAGTCTCTGAAACAGC 821
DB 10 GACATTGATAGTCTCTGAAACAGC 34
RESULT 37
ID ABRK33686 standard; DNA; 34 BP.
XX ABRK33686;
AC ABRK33686;
XX 08-MAY-2002 (first entry)
DT
XX S. pneumoniae BVH-11 gene, PCR primer HAMJ 284.
DE
XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; PCR; primer; ss.
XX Streptococcus pneumoniae.
OS
XX MO200198334-A2.
EN
XX 27-DEC-2001.
PD
XX 19-JUN-2001; 2001WO-CA00908.
PF
XX 20-JUN-2000; 2000US-212683P.
PR
XX (SHIR-) SHIRE BIOCHEM INC.
XX
XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
PI
XX WPI; 2002-122272/16.
DR
XX
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia -
XX
XX Example 1; Page 31; 113pp; English.
PS
XX The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. norcardia or
CC Streptococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a primer

CC used for the isolation of S. pneumoniae genes from which the antigenic
CC peptides of the invention are derived.
XX
SQ Sequence 34 BP; 10 A; 8 C; 7 G; 9 T; 0 other;
Query Match 1.0%; Score 25; DB 24; Length 34;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 797 GACATTGATAGTCTCTGAAACAGC 821
DB 10 GACATTGATAGTCTCTGAAACAGC 34
RESULT 38
ID ABRK33719 standard; DNA; 35 BP.
XX ABRK33719;
AC ABRK33719;
XX 08-MAY-2002 (first entry)
DT
XX S. pneumoniae BVH-11 gene, PCR primer HAMJ 403.
DE
XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; PCR; primer; ss.
XX Streptococcus pneumoniae.
OS
XX MO200198334-A2.
EN
XX 27-DEC-2001.
PD
XX 19-JUN-2001; 2001WO-CA00908.
PF
XX 20-JUN-2000; 2000US-212683P.
PR
XX (SHIR-) SHIRE BIOCHEM INC.
XX
XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
PI
XX WPI; 2002-122272/16.
DR
XX
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia -
XX
XX Example 1; Page 33; 113pp; English.
PS
XX The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. norcardia or
CC Streptococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a primer
CC used for the isolation of S. pneumoniae genes from which the antigenic
CC peptides of the invention are derived.
XX
SQ Sequence 35 BP; 10 A; 8 C; 8 G; 9 T; 0 other;

Query Match 1.0%; Score 25; DB 24; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.056;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 797 GACATTGATAGTCTCTTGAACAGC 821
DB 11 GACATTGATAGTCTCTTGAACAGC 35

RESULT 39

AAV27414
ID AAV27414 standard; DNA; 1342 BP.

AC AAV27414;

DT 02-OCT-1998 (first entry)

DE Streptococcus pneumoniae SPI03 nucleotide.

KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
XX detection; pneumonia; otitis media; meningitis; ss.

OS Streptococcus pneumoniae.

XX Key Location/Qualifiers

FT CDS 2..1342

FT /tag= a

FT /product= "SPI03"

FT /note= "no stop codon given"

PN WO9818930-A2.

PD 07-MAY-1998.

PF 30-OCT-1997; 97WO-US19422.

PR 31-OCT-1996; 96US-0023960.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Choi GH, Hromockyj A, Johnson LS, Kunsch CA;

DR WPI: 1998-272224/24.

DR P-PSDB; AAW61228.

PT Nucleic acid encoding antigenic peptide(s) from Streptococcus
XX pneumoniae - or their epitope-containing fragments, useful in
XX protective or therapeutic vaccines, and for diagnosis

PS Claim 1; Page 85; 118pp; English.

CC The present sequence encodes a protein from Streptococcus pneumoniae.
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein
CC can be useful in vaccines for inducing protective antibodies against
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
CC are used to detect Streptococcus infection (by usual hybridisation or
CC amplification methods), also for isolating Streptococcus genes or their
CC allelic variants. The protein can be used similarly to detect specific
CC antibodies in standard immunoassays, especially for diagnosing or
CC monitoring infections. Antibodies which bind the protein are used to
CC detect corresponding antigens, to purify the protein and for passive
CC immunisation (optionally coupled to a toxin). Vaccines are administered,
CC e.g. by injection, orally or through the skin, typically at 0.01-1000
CC (especially 10-300) mu g/ml per dose.

CC Sequence 1342 BP; 451 A; 270 C; 261 G; 360 T; 0 other;

Query Match 1.0%; Score 24; DB 19; Length 1342;

Best Local Similarity 100.0%; Pred. No. 0.21;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 546 ATGCTTATATCGTCTCCTCATGAG 569

DB 525 ATGCTTATATCGTCTCCTCATGAG 548

RESULT 40

ABQ84882
ID ABQ84882 standard; DNA; 1342 BP.

AC ABQ84882;

DT 04-SEP-2002 (first entry)

DE S. pneumoniae SPI03 nucleotide sequence SEQ ID NO:181.

KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
XX antibacterial; Streptococcal infection; detection; gene; ds.

OS Streptococcus pneumoniae.

PN US2002061545-A1.

PD 23-MAY-2002.

PF 22-JAN-2001; 2001US-0765272.

PR 30-OCT-1997; 97US-0961083.

PA (CHOI/) CHOI G H.

PA (KUNS/) KUNSCH C A.

PA (BARA/) BARASH S C.

PA (DILL/) DILLON P J.

PA (DOUG/) DOUGHERTY B.

PA (FANN/) FANNON M R.

PA (ROSE/) ROSEN C A.

PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;

DR WPI: 2002-479261/51.

DR P-PSDB; ABP54647.

PT New Streptococcus pneumoniae antigens, useful for detecting
XX Streptococcus and for preventing or attenuating disease caused by
XX Streptococcus infection -

PS Claim 1; Page 45; 70pp; English.

CC ABQ84792 to ABQ84904 represents nucleic acids which encode the
CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669.
CC The S. pneumoniae antigens have antibacterial activity and can be
CC used in vaccines. The S. pneumoniae antigens can also be used to
CC prevent or attenuate a Streptococcal infection in an animal. The
CC polynucleotides encoding the S. pneumoniae antigens can be used to
CC detect Streptococcus nucleic acids. ABQ84905 to ABQ85130 represent
CC primers used in the cloning of S. pneumoniae ORFs (open reading frames)
XX which are used in an example from the present invention.

CC Sequence 1342 BP; 451 A; 270 C; 261 G; 360 T; 0 other;

Query Match 1.0%; Score 24; DB 24; Length 1342;

Best Local Similarity 100.0%; Pred. No. 0.21;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 546 ATGCTTATATCGTCTCCTCATGAG 569

DB 525 ATGCTTATATCGTCTCCTCATGAG 548

RESULT 41

ABN66839
ID ABN66839 standard; DNA; 1398 BP.

XX ABN66839;

AC ABN66839;

```

XX 01-JUL-2002 (first entry)
XX Streptococcus polynucleotide SEQ ID NO 1591.
DE Streptococcus polynucleotide SEQ ID NO 1591.
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
XX antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX Streptococcus agalactiae.
XX WO200234771-A2.
XX 02-MAY-2002.
XX 29-OCT-2001; 2001WO-GB04789.
XX 27-OCT-2000; 2000GB-0026333.
XX 24-NOV-2000; 2000GB-0028727.
XX 07-MAR-2001; 2001GB-0005640.
XX (CHIR-) CHIRON SPA.
XX (GENO-) INST GENOMIC RES.
XX Telford J, Maignani V, Margalit Ros YI, Grandi G, Fraser C;
XX Tetelín H;
XX W21, 2002-352536/38.
XX P-PSDB; ABP26208.
XX New Streptococcus protein for the treatment or prevention of infection
XX or disease caused by Streptococcus bacteria, such as meningitis, and
XX for detecting a compound that binds to the protein -
XX Claim 7, Page 3309-3310; 4525pp; English.
XX The invention relates to a protein (ABP25413-ABP30895) from group B
XX Streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
XX (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
XX the specification. The proteins have antibacterial and antiinflammatory
XX activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and
XX antibodies that bind (I) are used in the manufacture of medicaments for
XX the treatment or prevention of infection or disease caused by
XX Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
XX Nucleic acids encoding (I) are used to detect Streptococcus in a
XX biological sample. (I) is used to determine whether a compound binds to
XX (I). A composition comprising (I) or a nucleic acid encoding (I), may be
XX used as a vaccine or diagnostic composition. The disease caused by
XX Streptococcus that is prevented or treated may be meningitis. Nucleic
XX acid encoding (I) may be used to recombinantly produce (I) and may be
XX used in gene therapy. Antibodies to (I) are used for affinity
XX chromatography, immunoassays, and distinguishing/identifying
XX Streptococcus proteins.
XX Sequence 1398 BP; 502 A; 273 C; 271 G; 352 T; 0 other;
SQ
Query March 1.0%; Score 24; DB 24; Length 1398;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 485 CGCTATACAGATGATGTTAT 508
DB 583 CGCTATACAGATGATGTTAT 606

```

```

DE Streptococcus pneumoniae BVH-3A gene SEQ ID NO:7.
XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
XX prophylaxis; therapy; infection; diagnosis; meningitis; bacteremia;
XX otitis media; pneumonia; immunisation; bactericidal; ds.
XX Streptococcus pneumoniae.
XX WO200039299-A2.
XX 06-JUL-2000.
XX 20-DEC-1999; 99WO-CAN01218.
XX 23-DEC-1998; 98US-0113800.
XX (BIOC-) BIOCHEM PHARMA INC.
XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI: 2000-452397/39.
XX P-PSDB; AAB12718.
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX otitis media, bacteremia and/or pneumonia -
XX Disclosure; Fig 7; 106pp; English.
XX The present invention describes nucleic acids (I) encoding protein
XX antigens (II) from Streptococcus pneumoniae. The protein antigens
XX have bactericidal activity. The nucleic acids, encoding the proteins
XX antigens, may be used for the recombinant production of the proteins
XX they encode. The protein antigens may then be used as vaccines for the
XX prevention and treatment of Streptococcal infections in mammals
XX (especially humans) which result in, e.g. meningitis, otitis media,
XX bacteraemia and/or pneumonia. The present sequence encodes the
XX S. pneumoniae BVH-3A protein antigen.
XX Sequence 1455 BP; 491 A; 286 C; 285 G; 393 T; 0 other;
SQ
Query March 1.0%; Score 24; DB 21; Length 1455;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 546 ATGCTATATGCTTCCTCATGAG 569
DB 596 ATGCTATATGCTTCCTCATGAG 619

```

```

RESULT 43
ID AAA47603 standard; DNA; 1455 BP.
AC AAA47603;
DT 20-OCT-2000 (first entry)
DE Recombinant variant of Sp36 gene (Sp36E) of S. pneumoniae.
XX Streptococcus pneumoniae; infection; vaccine; coiled coil region;
XX histidine triad residue; Sp36; antibody; otitis media;
XX nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
XX meningitis; lobar pneumonia; ds.
XX Streptococcus pneumoniae.
OS Streptococcus pneumoniae.
XX
XX Key Location/Qualifiers
FH 1..1455
FT CDS /*tag= a
FT /product= Sp36E polypeptide
FT
FT
PV WO200037105-A2.
XX

```



```

PD 29-JUN-2000.
XX
XX 21-DEC-1999; 99WO-US30390.
XX
XX 21-DEC-1998; 98US-0113048.
XX
XX (MEDI-) MEDIMUNE INC.
XX
XX Johnson LS, Koenig S, Adamou JE;
XX
XX MPI; 2000-452129/39.
XX
XX P-PSDB; AAB01467.
XX
XX Vaccine useful for prophylaxis and treatment of pneumococcal infections
XX such as otitis media, nasopharyngeal and bronchial infections,
XX comprises Streptococcus pneumoniae proteins
XX
XX Disclosure; Page 60-61; 70pp; English.
XX
XX Although a number of proteins have been suggested as being involved
XX in the pathogenicity of Streptococcus pneumoniae, there still remains
XX a need to identify polypeptides having epitopes in common from
XX various strains of S. pneumoniae in order to utilize such
XX polypeptides in vaccines to protect against a wide variety of
XX S. pneumoniae. New vaccine compositions are described which comprise a
XX Streptococcus pneumoniae polypeptide (or fragments) of 80 - 680 amino
XX acids in length that comprise at least one histidine triad residue
XX (HxxHxx) or a coiled-coil region, or an antibody directed against
XX these features. The vaccine is useful in protecting against infection
XX by Streptococcus pneumoniae. The vaccine composition comprising
XX antibodies to is useful for passive immunization for treating
XX pneumococcal infections which includes otitis media, nasopharyngeal
XX and bronchial infections.
XX
SQ Sequence 1455 BP; 491 A; 287 C; 285 G; 392 T; 0 other;

Query Match          1.0%; Score 24; DB 21; Length 1455;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 546 ATGCTTATATCGTTCTCATGAG 569
DB 596 ATGCTTATATCGTTCTCATGAG 619

RESULT 44
AAA05473
ID AAA05473 standard; DNA; 1455 BP.
XX
XX AAA05473;
XX
XX 24-MAY-2000 (first entry)
XX
XX Streptococcus pneumoniae type 4 nucleotide sequence 4138.1.
XX
XX Streptococcus pneumoniae; vaccine; screening; protein antigen;
XX antibacterial; antiinflammatory; meningitis; infection; diagnosis;
XX pneumococcal disease; ds.
XX
XX Streptococcus pneumoniae.
XX
XX WO200006737-A2.
XX
XX 10-FEB-2000.
XX
XX 27-JUL-1999; 99WO-GB02451.
XX
XX 27-JUL-1998; 98GB-0016337.
XX
XX 19-MAR-1999; 99US-0125164.
XX
XX (MICR-) MICROBIAL TECHNIQS LTD.
XX
XX Gilbert CFG, Hansbro PM;
XX
PI

```

```

XX
XX MPI; 2000-195300/17.
XX
XX New Streptococcal protein, useful as a vaccine, for diagnosis of
XX pneumococcal diseases and for screening agents capable of antagonizing
XX or inhibiting expression of the protein
XX
XX Claim 7; Page 41-42; 108pp; English.
XX
XX AA081501 to AA081679 represent specifically claimed protein sequences
XX isolated from Streptococcus pneumoniae. AA05407 to AA05390 represent
XX specifically claimed nucleotide sequences isolated from S. pneumoniae.
XX The sequences have antibacterial and antiinflammatory properties.
XX The protein sequences, and fragments of them, are useful as immunogens
XX and/or antigens. The nucleotide sequences can be used in vaccines and in
XX diagnostic assays. The proteins and nucleotides can be useful for the
XX detection and diagnosis of S. pneumoniae. The protein sequences are also
XX useful for screening an agent capable of antagonizing, inhibiting or
XX interfering with the function or expression of the proteins in which the
XX agent is useful for treatment or prophylaxis of S. pneumoniae infection
XX and meningitis. AA05591 to AA05614 represent primers used in the
XX exemplification of the present invention.
XX
SQ Sequence 1455 BP; 491 A; 287 C; 285 G; 392 T; 0 other;

Query Match          1.0%; Score 24; DB 21; Length 1455;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 546 ATGCTTATATCGTTCTCATGAG 569
DB 596 ATGCTTATATCGTTCTCATGAG 619

RESULT 45
AAZ91804
ID AAZ91804 standard; DNA; 1455 BP.
XX
XX AAZ91804;
XX
XX 02-JUN-2000 (first entry)
XX
XX Streptococcus pneumoniae DNA sequence ID128.
XX
XX Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;
XX bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;
XX kidney disease; diabetes; immunosuppressive disorder; otitis media;
XX pneumococcal septicaemia; sinusitis; meningitis; therapy; ss.
XX
XX Streptococcus pneumoniae.
XX
XX WO200006738-A2.
XX
XX 10-FEB-2000.
XX
XX 27-JUL-1999; 99WO-GB02452.
XX
XX 27-JUL-1998; 98GB-0016336.
XX
XX 19-MAR-1999; 99US-0125329.
XX
XX (MICR-) MICROBIAL TECHNIQS LTD.
XX
XX Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;
XX
XX MPI; 2000-195301/17.
XX
XX P-PSDB; AAY81708.
XX
XX Streptococcal proteins and polynucleotides useful for diagnosis,
XX treatment and prophylaxis of bacterial infections
XX
XX Claim 1; Page 39; 76pp; English.
XX
XX This sequence encodes a Streptococcus pneumoniae protein of the
XX
XX

```

CC invention. The proteins (or their homologues, derivatives and/or
 CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic
 CC compositions comprising the proteins are useful as vaccines and also in
 CC diagnostic assays. The sequences are useful for the detection or
 CC diagnosis of *S. pneumoniae* infection, by contacting a sample to be tested
 CC with them. Agents capable of antagonising, inhibiting or interfering with
 CC the function or expression of the protein or polypeptide are useful in
 CC medical compositions in the treatment or prophylaxis of *S. pneumoniae*
 CC infection. As the sequences can be used to treat *S. pneumoniae* infection,
 CC they can be used to treat bacterial pneumonia, which has high rates in
 CC young children, the elderly, and in patients with predisposing conditions
 CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,
 CC or with immunosuppressive disorders, especially AIDS. They can also be
 CC used to treat pneumococcal septicaemia, otitis media, sinusitis, and
 CC meningitis.

SO Sequence 1455 BP; 491 A; 287 C; 285 G; 392 T; 0 other;

Query Match 1.0%; Score 24; DB 21; Length 1455;
 Best Local Similarity 100.0%; Pred. No. 0.21;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 546 ATGCTTATATGTTCTCATGAG 569
 DB 596 ATGCTTATATGTTCTCATGAG 619

RESULT 46
 AAA65738
 ID AAA65738 standard; DNA; 2528 BP.
 AC AAA65738;
 XX
 XX 21-NOV-2000 (first entry)
 DT
 XX Streptococcus pneumoniae SP63 BVH-3 gene SEQ ID NO:15.
 DE
 XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KM otitis media; pneumonia; immunisation; bactericidal; ds.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN MO200039299-A2.
 XX
 PD 06-JUL-2000.
 XX
 XX 20-DEC-1999; 99WO-CA01218.
 PF
 XX 23-DEC-1998; 98US-0113800.
 PR
 XX (BIOC-) BIOCHEM PHARMA INC.
 PA
 P1 Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 XX
 DR MPI: 2000-452397/39.
 DR P-PSDB; AAB12721.
 DR
 PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteraemia and/or pneumonia -
 PT
 XX
 PS Example 6; Fig 18; 106pp; English.
 XX
 CC The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the protein
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence encodes the
 CC *S. pneumoniae* SP63 BVH-3 protein antigen.

SO Sequence 2528 BP; 960 A; 398 C; 505 G; 665 T; 0 other;

Query Match 1.0%; Score 24; DB 21; Length 2528;
 Best Local Similarity 100.0%; Pred. No. 0.21;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 546 ATGCTTATATGTTCTCATGAG 569
 DB 536 ATGCTTATATGTTCTCATGAG 559

RESULT 47
 ABRK15105
 ID ABRK15105 standard; cDNA; 2528 BP.
 XX
 AC ABRK15105;
 XX
 DT 08-MAY-2002 (first entry)
 DT
 XX DNA encoding Streptococcus pneumoniae BVH-3 version #2.
 DE
 XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 KM pneumonia; streptococcal bacterial infection; gene; ss.
 XX
 OS Streptococcus pneumoniae.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..2523
 FT /*tag= a
 FT /product= "BVH-3"
 FT
 PN MO200198334-A2.
 XX
 PD 27-DEC-2001.
 PD
 XX 19-JUN-2001; 2001WO-CA00908.
 PF
 XX 20-JUN-2000; 2000US-212683P.
 PR
 XX (SHIR-) SHIRE BIOCHEM INC.
 PA
 P1 Hamel J, Quellet C, Charland N, Martin D, Brodeur B;
 XX
 DR MPI: 2002-122272/16.
 DR
 DR P-PSDB; AAU76151.
 DR
 XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia -
 PT
 XX
 PS Example 8; Fig 9; 113pp; English.
 XX
 CC The invention describes an isolated polypeptide (I) with 70-90%
 CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic
 CC treatment of meningitis, otitis media, bacteraemia or pneumonia infection
 CC in an individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, *S. dysgalactiae*, *S. uberis*, *S. novae* or
 CC *Staphylococcus aureus*) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for *S. pneumoniae* infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating *S. pneumonia* nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence encodes the
 CC Streptococcus pneumoniae protein BVH-3, used to create the antigenic
 CC peptides described in the method of the invention.

XX. Sequence 2528 BP; 960 A; 398 C; 505 G; 665 T; 0 other;
SQ Query Match 1.0%; Score 24; DB 24; Length 2528;
Best Local Similarity 100.0%; Pred.No.0.21;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
546 ATGCTTATATCGTTCCTCATGAG 569
|||||
536 ATGCTTATATCGTTCCTCATGAG 559

Db

RESULT 48
ABX06706
ID ABX06706 standard; DNA; 3117 BP.
XX
XX ABX06706;
XX
XX 11-FEB-2003 (first entry)
XX
XX S. pneumoniae type 4 strain coding region #994.
DE
XX Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media;
KM ear infection; antiinflammatory; antibacterial; immunostimulant;
KW auditory; respiratory; gene therapy; vaccine.
XX
XX Streptococcus pneumoniae type 4 strain.
OS
XX WO200277021-A2.
XX
XX 03-OCT-2002.
XX
XX 27-MAR-2002; 2002WO-1B02163.
XX
XX 27-MAR-2001; 2001GB-0007658.
PR
XX (CHIR-) CHIRON SPA.
XX (GENO-) INST GENOMIC RES.
XX
XX Masignani V, Tetteijn H; Fraser C;
PI WPI; 2003-040579/03.
XX
XX P-PSDB; ABU01419.
XX
XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT useful as medicaments for treating or preventing a disease or infection
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
PT or ear infection
XX
XX Claim 6; SEQ ID No 1987; 56pp; English.

XX
XX The invention relates to a protein comprising or having at least 50%
CC identity to any of the 2469 amino acid sequences, identified in the
CC specification (available on a computer readable format), or its fragment,
CC expressed from 2469 of 2489 identified DNA coding regions from the
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC ABS56454. Also included are an antibody which binds one of the
CC proteins, treating a patient by administering the protein, DNA or
CC antibody (in a composition), a kit comprising first and second primers,
CC which are the nucleic acid cited above or fragments between nucleotides
CC 8-100 of a sequence not defined in the specification, for amplifying a
CC target sequence contained within a Streptococcus nucleic acid sequence,
CC where the first primer is substantially complementary to the target
CC sequence and the second primer is substantially complementary to the
CC complement of the target sequence, and where the parts of the primers
CC having substantial complementarity define the termini of the target
CC sequence to be amplified, assay comprising contacting a test compound
CC with the protein, and determining whether the test compound binds to the
CC protein and a Streptococcus pneumoniae bacterium, where one or more
CC genes encoding the proteins has been rendered inactive. The proteins,
CC nucleic acid molecules, antibody and compositions are useful as
CC medicaments for treating or preventing a disease or infection due to
CC streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,

CC sepsis, otitis media or ear infection. They are also useful in developing
CC vaccines, diagnostics and antibiotics. The methods are useful for
CC identifying immunodominant proteins. The present sequence is one of
CC the 2489 identified coding region from the genomic sequence.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3117 BP; 1150 A; 526 C; 612 G; 829 T; 0 other;
Query Match 1.0%; Score 24; DB 25; Length 3117;
Best Local Similarity 100.0%; Pred.No.0.21;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
546 ATGCTTATATCGTTCCTCATGAG 569
|||||
596 ATGCTTATATCGTTCCTCATGAG 619

Db

RESULT 49
AAA65730
ID AAA65730 standard; DNA; 3120 BP.
XX
XX AAA65730;
XX
XX 21-NOV-2000 (first entry)
XX
XX Streptococcus pneumoniae BVH-3 gene SEQ ID NO:1.
DE
XX Streptococcus pneumoniae BVH-3; BVH-11; BVH-28; antigen; vaccine;
XX prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KM otitis media; pneumonia; immunisation; bactericidal; ds.
XX
XX Streptococcus pneumoniae.
OS
XX WO200039299-A2.
XX
XX 06-JUL-2000.
XX
XX 20-DEC-1999; 99WO-CA01218.
XX
XX 23-DEC-1998; 98US-0113800.
XX
XX (BIOC-) BIOCHEM PHARMA INC.
XX
XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
PI WPI; 2000-452397/39.
XX
XX P-PSDB; AAB12715.
XX
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia
PT
XX Example 1; Fig 1; 106pp; English.

XX
XX The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence encodes the
CC S. pneumoniae BVH-3 protein antigen.
XX
SQ Sequence 3120 BP; 1152 A; 525 C; 612 G; 831 T; 0 other;
Query Match 1.0%; Score 24; DB 21; Length 3120;
Best Local Similarity 100.0%; Pred.No.0.21;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
546 ATGCTTATATCGTTCCTCATGAG 569

```

Db      596 ATGCTTATATCGTTCCTCATGAG 619
|||||
RESULT 50
AAA65735
ID      AAA65735 standard; DNA; 5048 BP.
XX
XX      AAA65735;
AC
XX      21-NOV-2000 (first entry)
DT
XX      Streptococcus pneumoniae BVH-3 gene SEQ ID NO:11.
DE
XX      Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
XX      prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW      otitis media; pneumonia; immunisation; bactericidal; ds.
XX
XX      Streptococcus pneumoniae.
OS
XX      WO200039299-A2.
XX      06-JUL-2000.
XX
XX      20-DEC-1999; 99WO-CA01218.
XX      23-DEC-1998; 98US-0113800.
XX      (BIOC-) BIOCHEM PHARMA INC.
XX      Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX      WPI, 2000-452397/39.
XX      Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX      otitis media, bacteraemia and/or pneumonia -
XX      Example 6; Fig 14; 106pp; English.
XX
XX      The present invention describes nucleic acids (I) encoding protein
XX      antigens (II) from Streptococcus pneumoniae. The protein antigens
XX      have bactericidal activity. The nucleic acids, encoding the protein
XX      antigens, may be used for the recombinant production of the proteins
XX      they encode. The protein antigens may then be used as vaccines for the
XX      prevention and treatment of Streptococcal infections in mammals
XX      (especially humans) which result in, e.g. meningitis, otitis media,
XX      bacteraemia and/or pneumonia. The present sequence encodes the
XX      S. pneumoniae BVH-3 protein antigen.
XX
XX      Sequence 5048 BP; 1709 A; 907 C; 1104 G; 1328 T; 0 other;
SQ
Query Match      1.0%; Score 24; DB 21; Length 5048;
Best Local Similarity 100.0%; Pred. No. 0.22;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      546 ATGCTTATATCGTTCCTCATGAG 569
Db      2372 ATGCTTATATCGTTCCTCATGAG 2395
|||||
RESULT 51
ABK15101
ID      ABK15101 standard; DNA; 5048 BP.
XX
XX      ABK15101;
AC
XX      08-MAY-2002 (first entry)
DT
XX      DNA encoding Streptococcus pneumoniae BVH-3 version #1.
DE
XX      BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW      pneumonia; streptococcal bacterial infection; gene; ds.
XX

```

```

OS      Streptococcus pneumoniae.
XX
XX      Key      Location/Qualifiers
XX      CDS      1777..4896
XX              /tag=a
XX              /product="BVH-3"
XX              /note="The gene is flanked by sequences from the
XX              vector Sp64, no information on which is
XX              given in the specification"
XX
XX      WO200198334-A2.
XX
XX      27-DEC-2001.
XX
XX      19-JUN-2001; 2001WO-CA09098.
XX      20-JUN-2000; 2000US-212683P.
XX      (SHIR-) SHIRE BIOCHEM INC.
XX      Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX      WPI, 2002-122272/16.
XX      P-PSDB; AAU75932.
XX
XX      New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX      epitope-bearing polypeptides, useful as vaccine components for treating
XX      or preventing streptococcal infections such as otitis media,
XX      meningitis, and bacteraemia -
XX      Example 3; Fig 2; 113pp; English.
XX
XX      The invention describes an isolated polypeptide (I) with 70-90%
XX      identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
XX      BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
XX      comprising (I) is useful for therapeutic or prophylactic treatment of
XX      meningitis, otitis media, bacteraemia or pneumonia infection in an
XX      individual susceptible to these disorders. (II) is also useful for
XX      therapeutic or prophylactic treatment of any streptococcal bacterial
XX      infection (e.g., caused by Streptococcus pneumoniae, group A
XX      Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
XX      as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novacida or
XX      Streptococcus aureus) in an individual susceptible to the infection.
XX      A polynucleotide (III) encoding (I) is useful in DNA immunisation
XX      techniques. The Streptococcus polypeptides are useful in a diagnostic
XX      test for S. pneumoniae infection. (III) is useful for designing DNA
XX      probes for use in detecting the presence of Streptococcus in a biological
XX      sample suspected of containing the bacteria. The DNA probes may also be
XX      used for detecting circulating S. pneumonia nucleic acid in a sample for
XX      diagnosing streptococcal infections. This sequence encodes the
XX      Streptococcus pneumoniae protein BVH-3, used to create the antigenic
XX      peptides described in the method of the invention.
XX
XX      Sequence 5048 BP; 1709 A; 907 C; 1104 G; 1328 T; 0 other;
SQ
Query Match      1.0%; Score 24; DB 24; Length 5048;
Best Local Similarity 100.0%; Pred. No. 0.22;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      546 ATGCTTATATCGTTCCTCATGAG 569
Db      2372 ATGCTTATATCGTTCCTCATGAG 2395
|||||
RESULT 52
AAV52325/C
ID      AAV52325 standard; DNA; 6867 BP.
XX
XX      AAV52325;
AC
XX      23-OCT-1998 (first entry)
DT
XX      Streptococcus pneumoniae genome fragment SEQ ID NO:192.
DE

```

XX Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
 KW computer readable medium; vaccine; pharmaceutical composition; ds.
 XX
 OS Streptococcus pneumoniae.

XX WO9818931-A2.

XX PD 07-MAY-1998.

XX PF 30-OCT-1997; 97WO-US19588.

XX PR 31-OCT-1996; 96US-0029960.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,
 XX Kunsch CA, Rosen CA,
 XX WPI, 1998-272225/24.

PT Computer-readable medium with recorded Streptococcus pneumoniae
 PT polynucleotide sequences - useful in diagnostic kits and assays, and
 PT pharmaceutical compositions and vaccines for Streptococcus
 PS pneumoniae

XX Claim 1, Page 1157-1161; 1409pp; English.

XX The present invention describes a computer readable medium which has
 CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
 CC recorded on it, or a representative fragment or a sequence at least 95%
 CC identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in
 CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
 CC Streptococcus pneumoniae. The present invention also describes an
 CC isolated nucleic acid molecule encoding a homologue of any of the
 CC fragments of the S. pneumoniae genome (SEQ ID NO:1 to 391) where the
 CC nucleic acid molecule is produced by a process comprising: (a) screening
 CC a genomic DNA library using as a probe a target sequence defined by any
 CC of the sequences in SEQ ID NO:1 to 391, identifying members of the
 CC library which contain sequences that hybridise to the target sequence and
 CC isolating the nucleic acid molecules from the members; or (b) isolating
 CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
 CC molecules whose nucleotide sequence is homologous to amplification
 CC primers derived from the fragment of the S. pneumoniae genome to prime
 CC the amplification and isolating the amplified sequences. The computer
 CC readable medium can be used in a computer-based system for identifying
 CC fragments of the S. pneumoniae genome of commercial importance, or
 CC expression modulating fragments of the S. pneumoniae genome. Products
 CC from the present invention can be used in diagnosis kits and assays, and
 CC pharmaceutical compositions and vaccines for S. pneumoniae.

XX Sequence 6867 BP; 1896 A; 1325 C; 1212 G; 2433 T; 1 other;

Query Match 1.0%; Score 24; DB 19; Length 6867;
 Best Local Similarity 100.0%; Pred. No. 0.22;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 546 ATGCTTATTCGTTCTCATGGAG 569

DB 6190 ATGCTTATTCGTTCTCATGGAG 6167

RESULT 53

ABN71527/c
 ID ABN71527 standard; DNA; 2155561 BP.

XX ABN71527;

DT 02-JUL-2002 (first entry)

XX Streptococcus polynucleotide SEQ ID NO 10967.

XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;

KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
 XX
 OS Streptococcus sp.

XX WO200234771-A2.

XX PD 02-MAY-2002.

XX PF 29-OCT-2001; 2001WO-GB04789.

XX PR 27-OCT-2000; 2000GB-0026333.

XX PR 24-NOV-2000; 2000GB-0028727.

XX PR 07-MAR-2001; 2001GB-0005640.

XX PA (CHIR-) CHIRON SPA.

XX PA (GENO-) INST GENOMIC RES.

XX PI Telford J, Maignani V, Margarit Ros YI, Grandi G, Fraser C;
 XX Tettelein H;
 XX WPI, 2002-352536/38.

PT New Streptococcus protein for the treatment or prevention of infection
 PT or disease caused by Streptococcus bacteria, such as meningitis, and
 PT for detecting a compound that binds to the protein -

XX Claim 8, Page 4196-4488; 4525pp; English.

XX The invention relates to a protein (ABP25413-ABP30895) from group B
 CC Streptococcus pyogenes (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (1), nucleic acids encoding (1), ABN6044-ABN71526 and
 CC antibodies that bind (1) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (1) are used to detect Streptococcus in a
 CC biological sample. (1) is used to determine whether a compound binds to
 CC (1). A composition comprising (1) or a nucleic acid encoding (1), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (1) may be used to recombinantly produce (1) and may be
 CC used in gene therapy. Antibodies to (1) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.

XX Sequence 2155561 BP; 695741 A; 385320 C; 382958 G; 691542 T; 0 other;

Query Match 1.0%; Score 24; DB 24; Length 2155561;
 Best Local Similarity 100.0%; Pred. No. 0.27;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 485 CGCTTACTACAGATGATGTTAT 508

DB 1890709 CGCTTACTACAGATGATGTTAT 1890686

RESULT 54
 AAC37581
 ID AAC37581 standard; DNA; 1547 BP.

XX AAC37581;

DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 17910.

XX Hybridisation assay; genetic mapping; gene expression control;

XX protein identification; signal transduction pathway;

XX metabolic pathway; promoter; termination sequence; ss.
 OS Arabidopsis thaliana.

XX	EP1033405-A2.		PR	01-JUL-1999;	99US-0141842.
PN			PR	01-JUL-1999;	99US-0142154.
XX	06-SEP-2000.		PR	02-JUL-1999;	99US-0142055.
PD			PR	06-JUL-1999;	99US-0142390.
XX			PR	08-JUL-1999;	99US-0142803.
PF	25-FEB-2000; 2000EP-0301439.		PR	09-JUL-1999;	99US-0142920.
XX			PR	12-JUL-1999;	99US-0142977.
XX			PR	13-JUL-1999;	99US-0143542.
PR	25-FEB-1999;	99US-0121825.	PR	14-JUL-1999;	99US-0143624.
PR	05-MAR-1999;	99US-0123180.	PR	15-JUL-1999;	99US-0144005.
PR	09-MAR-1999;	99US-0123548.	PR	16-JUL-1999;	99US-0144085.
PR	23-MAR-1999;	99US-0125788.	PR	19-JUL-1999;	99US-0144325.
PR	25-MAR-1999;	99US-0126264.	PR	19-JUL-1999;	99US-0144332.
PR	29-MAR-1999;	99US-0126785.	PR	19-JUL-1999;	99US-0144334.
PR	01-APR-1999;	99US-0127462.	PR	19-JUL-1999;	99US-0144334.
PR	06-APR-1999;	99US-0128234.	PR	19-JUL-1999;	99US-0144334.
PR	08-APR-1999;	99US-0128714.	PR	19-JUL-1999;	99US-0144334.
PR	16-APR-1999;	99US-0128845.	PR	19-JUL-1999;	99US-0144334.
PR	19-APR-1999;	99US-0130077.	PR	20-JUL-1999;	99US-0144332.
PR	21-APR-1999;	99US-0130449.	PR	20-JUL-1999;	99US-0144332.
PR	23-APR-1999;	99US-0130510.	PR	20-JUL-1999;	99US-0144332.
PR	23-APR-1999;	99US-0130891.	PR	20-JUL-1999;	99US-0144332.
PR	28-APR-1999;	99US-0131449.	PR	21-JUL-1999;	99US-0144814.
PR	30-APR-1999;	99US-0132048.	PR	21-JUL-1999;	99US-0145086.
PR	30-APR-1999;	99US-0132407.	PR	21-JUL-1999;	99US-0145088.
PR	04-MAY-1999;	99US-0132484.	PR	22-JUL-1999;	99US-0145087.
PR	05-MAY-1999;	99US-0132485.	PR	22-JUL-1999;	99US-0145089.
PR	06-MAY-1999;	99US-0132486.	PR	22-JUL-1999;	99US-0145192.
PR	06-MAY-1999;	99US-0132487.	PR	23-JUL-1999;	99US-0145145.
PR	07-MAY-1999;	99US-0132863.	PR	23-JUL-1999;	99US-0145218.
PR	11-MAY-1999;	99US-0134256.	PR	23-JUL-1999;	99US-0145224.
PR	14-MAY-1999;	99US-0134218.	PR	26-JUL-1999;	99US-0145276.
PR	14-MAY-1999;	99US-0134219.	PR	27-JUL-1999;	99US-0145913.
PR	14-MAY-1999;	99US-0134370.	PR	27-JUL-1999;	99US-0145918.
PR	18-MAY-1999;	99US-0134768.	PR	27-JUL-1999;	99US-0145919.
PR	19-MAY-1999;	99US-0134941.	PR	28-JUL-1999;	99US-0145951.
PR	20-MAY-1999;	99US-0135124.	PR	02-AUG-1999;	99US-0146386.
PR	21-MAY-1999;	99US-0135353.	PR	02-AUG-1999;	99US-0146388.
PR	24-MAY-1999;	99US-0135629.	PR	02-AUG-1999;	99US-0146389.
PR	25-MAY-1999;	99US-0136021.	PR	03-AUG-1999;	99US-0147038.
PR	27-MAY-1999;	99US-0136392.	PR	04-AUG-1999;	99US-0147204.
PR	28-MAY-1999;	99US-0136782.	PR	04-AUG-1999;	99US-0147302.
PR	01-JUN-1999;	99US-0137222.	PR	05-AUG-1999;	99US-0147192.
PR	03-JUN-1999;	99US-0137528.	PR	05-AUG-1999;	99US-0147260.
PR	04-JUN-1999;	99US-0137502.	PR	06-AUG-1999;	99US-0147303.
PR	07-JUN-1999;	99US-0137724.	PR	06-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	99US-0138094.	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	99US-0138540.	PR	09-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	99US-0138847.	PR	10-AUG-1999;	99US-0148171.
PR	14-JUN-1999;	99US-0139119.	PR	11-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	99US-0139452.	PR	12-AUG-1999;	99US-0148341.
PR	17-JUN-1999;	99US-0139453.	PR	13-AUG-1999;	99US-0148555.
PR	17-JUN-1999;	99US-0139492.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139454.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139455.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139456.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139457.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139462.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151066.
PR	21-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151080.
PR	22-JUN-1999;	99US-0139899.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	99US-0140354.	PR	01-SEP-1999;	99US-0151930.
PR	24-JUN-1999;	99US-0140695.	PR	07-SEP-1999;	99US-0152353.
PR	28-JUN-1999;	99US-0140823.	PR	10-SEP-1999;	99US-0153070.
PR	29-JUN-1999;	99US-0140991.	PR	13-SEP-1999;	99US-0153758.
PR	30-JUN-1999;	99US-0141287.	PR	15-SEP-1999;	99US-0154018.

```

PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160880.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 0.9%; Score 22; DB 21; Length 1547;
 Best Local Similarity 100.0%; Pred. No. 2.1; Mismatches 0; Indels 0; Gaps 0;

```

QY 2285 ACTCTTCAATTATGATTAACA 2306
DB 447 ACTCTTCAATTATGATTAACA 468

```

```

RESULT 55
ID AAA65761 standard; DNA; 29 BP.
AC AAA65761;
DT 21-NOV-2000 (first entry)
DE Streptococcus pneumoniae BVH-11 PCR primer SEQ ID NO:36.
XX
XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal; PCR primer; ss.
XX
XX Streptococcus pneumoniae.
OS
XX WO200039299-A2.
XX
XX 06-JUL-2000.
XX
XX 20-DEC-1999; 99WO-CA01218.
XX

```

```

PR 23-DEC-1998; 98US-0113800.
XX
XX (BIOC-) BIOCHEM PHARMA INC.
PA
XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
PI WPI; 2000-452397/39.
XX
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia
XX
XX Example 8; Page 49; 106pp; English.
PS
XX The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents a PCR
CC primer for a S. pneumoniae GAS BVH-71 protein antigen, which is used in
CC an example from the present invention.
XX
XX Sequence 29 BP; 7 A; 8 C; 6 G; 8 T; 0 other;
SQ

```

Query Match 0.9%; Score 21; DB 21; Length 29;
 Best Local Similarity 100.0%; Pred. No. 5.9; Mismatches 21; Conservative 0; Indels 0; Gaps 0;

```

QY 621 CTGCAGAGCCTTCTATCTG 641
DB 9 CTGCAGAGCCTTCTATCTG 29

```

```

RESULT 56
ID ABK33683 standard; DNA; 29 BP.
AC ABK33683;
DT 08-MAY-2002 (first entry)
XX
XX S. pneumoniae BVH-11 gene, PCR primer HAMJ 264.
DE
XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; PCR; primer; ss.
XX
XX Streptococcus pneumoniae.
OS
XX WO200198334-A2.
XX
XX 27-DEC-2001.
XX
XX 19-JUN-2001; 2001WO-CA00908.
XX
XX 20-JUN-2000; 2000US-212683P.
XX
XX (SHIR-) SHIRE BIOCHEM INC.
PA
PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX
XX WPI; 2002-122272/16.
XX
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
XX meningitis, and bacteraemia
XX
XX Example 1; Page 31; 113pp; English.
XX
XX The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of

```


CC diagnosis of Candida-mediated diseases (local or systemic infections);
CC and (ii) to screen for agents that inhibit expression of hyphae-specific
CC proteins, potential therapeutic agents. The products of the invention
CC have fungicidal activity. (i), or related proteins and antibodies, are
CC specific for the hyphal (virulent) form of Candida albicans, so allow
CC differentiation between virulent and avirulent strains. This sequence
CC encodes the Candida albicans hyphal-specific protein Cap18p isolated from
CC contig4-2069.

XX Sequence 501 BP; 205 A; 59 C; 76 G; 161 T; 0 other;

XX Query Match 0.8%; Score 20; DB 24; Length 501;

XX Best Local Similarity 100.0%; Pred. No. 21;

XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX Db 1512 CAATTCCTCAATTCAGTAT 1531

XX Db 450 CAATTCCTCAATTCAGTAT 431

XX RESULT 59

XX ID AAA05814 standard; DNA; 1146 BP.

XX AAA05814;

XX 30-MAY-2000 (first entry)

XX Group B Streptococcus protein encoding nucleotide sequence SEQ ID NO:12.

XX Group B Streptococcus; Streptococcus agalactiae; protein antigen;

XX vaccine; screening; immunogen; detection; diagnosis; infection;

XX antibody; antibody; antibacterial; ds.

XX Streptococcus agalactiae.

XX WO200006736-A2.

XX 10-FEB-2000.

XX 27-JUL-1999; 99WO-GB02444.

XX 27-JUL-1998; 98GB-0016335.

XX 19-MAR-1999; 99US-0125163.

XX (MICR-) MICROBIAL TECHNIQS LTD.

XX Le Page RWF, Wells JM, Hanniffy SB;

XX WPI; 2000-195299/17.

XX P-PSDB; AAY91286.

XX New Group B Streptococcus protein, useful as vaccine, for diagnosis of

XX Streptococcal infections and for screening of antibodies or antibodies

XX Claim 4; Fig 1; 123pp; English.

XX AAA05803 to AAA05872 encode proteins, polypeptides and peptides (given

XX in AAY91275 to AAY91343) isolated from Group B Streptococcus (GBS), also

XX known as Streptococcus agalactiae. The GBS polynucleotides and

XX polypeptides have antibacterial activity. Immunogenic compositions

XX comprising GBS polynucleotides or polypeptides can be used as vaccines

XX and for the treatment or prophylaxis of GBS infection. The

XX polynucleotides and polypeptides can also be used in the detection of GBS

XX and for screening DNA encoding bacterial cell envelope associated or

XX secreted antigens in gram positive bacteria. AAA05873 to AAA05941

XX represent primers used in the exemplification of the present invention.

XX Sequence 1146 BP; 389 A; 229 C; 234 G; 294 T; 0 other;

XX Query Match 0.8%; Score 20; DB 21; Length 1146;

XX Best Local Similarity 100.0%; Pred. No. 22;

XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 479 CAGGACGCTACTACTACAGA 498

XX 571 CAGGACGCTACTACTACAGA 590

XX Db 571 CAGGACGCTACTACTACAGA 590

XX RESULT 60

XX ID ABN69535 standard; DNA; 2466 BP.

XX ABN69535;

XX 01-JUL-2002 (first entry)

XX Streptococcus polynucleotide SEQ ID NO 6983.

XX Streptococcus; GAS; GBS; Group B streptococcus; Streptococcus agalactiae;

XX group A streptococcus; Streptococcus pyogenes; antibacterial; gene;

XX antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.

XX Streptococcus agalactiae.

XX WO200234771-A2.

XX 02-MAY-2002.

XX 29-OCT-2001; 2001WO-GB04789.

XX 27-OCT-2000; 2000GB-0026333.

XX 24-NOV-2000; 2000GB-0028727.

XX 07-MAR-2001; 2001GB-0005640.

XX (CHIR-) CHIRON SPA.

XX (GENO-) INST GENOMIC RES.

XX Telford J, Maignani V, Margarit Ros YI, Grandi G, Fraser C;

XX Tettelein H;

XX WPI; 2002-352536/38.

XX P-PSDB; ABP28904.

XX New Streptococcus protein for the treatment or prevention of infection

XX or disease caused by Streptococcus bacteria, such as meningitis, and

XX for detecting a compound that binds to the protein -

XX Claim 7; Page 3860-3861; 4525pp; English.

XX The invention relates to a protein (ABP25413-ABP30895) from group B

XX streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS

XX (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in

XX the specification. The proteins have antibacterial and antiinflammatory

XX activity. (i), nucleic acids encoding (i), ABN6044-ABN71526 and

XX antibodies that bind (i) are used in the manufacture of medicaments for

XX the treatment or prevention of infection or disease caused by

XX Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.

XX Nucleic acids encoding (i) are used to detect Streptococcus in a

XX biological sample. (i) is used to determine whether a compound binds to

XX (i). A composition comprising (i) or a nucleic acid encoding (i), may be

XX used as a vaccine or diagnostic composition. The disease caused by

XX Streptococcus that is prevented or treated may be meningitis. Nucleic

XX acid encoding (i) may be used to recombinantly produce (i) and may be

XX used in gene therapy. Antibodies to (i) are used for affinity

XX chromatography, immunoassays, and distinguishing/identifying

XX Streptococcus proteins.

XX Sequence 2466 BP; 841 A; 475 C; 513 G; 637 T; 0 other;

XX Query Match 0.8%; Score 20; DB 24; Length 2466;

XX Best Local Similarity 100.0%; Pred. No. 23;

XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 479 CAGGACGCTACTACTACAGA 498

XX |

Db 571 CAAGACGCTACTACTACAGA 590

RESULT 61

ID ABN70334 standard; DNA; 2466 BP.

AC ABN70334;

DT 01-JUL-2002 (first entry)

XX Streptococcus polynucleotide SEQ ID NO 8581.

DE Streptococcus polynucleotide SEQ ID NO 8581.

XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;

KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;

XX antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.

OS Streptococcus agalactiae.

XX WO200234771-A2.

XX 02-MAY-2002.

XX 29-OCT-2001; 2001WO-GB04789.

XX 27-OCT-2000; 2000GB-0026333.

XX 24-NOV-2000; 2000GB-0028727.

XX 07-MAR-2001; 2001GB-0005640.

XX (CHIR-) CHIRON SPA.

XX (GENO-) INST GENOMIC RES.

XX Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C,

PI Tettein H;

XX WPI; 2002-352536/38.

XX P-PSDB; ABP29703.

XX New Streptococcus protein for the treatment or prevention of infection

PT or disease caused by Streptococcus bacteria, such as meningitis, and

PT for detecting a compound that binds to the protein.

XX Claim 7; Page 3969; 4525pp; English.

XX The invention relates to a protein (ABP25413-ABP30895) from group B

CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS

CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in

CC the specification. The proteins have antibacterial and antiinflammatory

CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and

CC antibodies that bind (I) are used in the manufacture of medicaments for

CC the treatment or prevention of infection or disease caused by

CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.

CC Nucleic acids encoding (I) are used to detect Streptococcus in a

CC biological sample. (I) is used to determine whether a compound binds to

CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be

CC used as a vaccine or diagnostic composition. The disease caused by

CC Streptococcus that is prevented or treated may be meningitis. Nucleic

CC acid encoding (I) may be used to recombinantly produce (I) and may be

CC used in gene therapy. Antibodies to (I) are used for affinity

CC chromatography, immunoassays, and distinguishing/identifying

CC Streptococcus proteins.

XX Sequence 2466 BP; 841 A; 475 C; 513 G; 637 T; 0 other;

SO Query Match 0.8%; Score 20; DB 24; Length 2466;

Best Local Similarity 100.0%; Pred. No. 23;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 479 CAAGACGCTACTACTACAGA 498

DB 571 CAAGACGCTACTACTACAGA 590

RESULT 62

ID AAA65740

XX AAA65740 standard; DNA; 2469 BP.

AC AAA65740;

DT 21-NOV-2000 (first entry)

XX Streptococcus pneumoniae GBS BVH-71 gene SEQ ID NO:80.

DE Streptococcus pneumoniae GBS BVH-71 gene SEQ ID NO:80.

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;

XX otitis media; pneumonia; immunisation; bactericidal; ds.

OS Streptococcus pneumoniae.

XX WO200039299-A2.

XX 06-JUL-2000.

XX 20-DEC-1999; 99WO-CA01218.

XX 23-DEC-1998; 98US-0113800.

XX (BIOC-) BIOCHEM PHARMA INC.

XX Hanel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

PI WPI; 2000-452397/39.

XX P-PSDB; AAB12746.

XX Streptococcal antigens useful for vaccinating against e.g. meningitis,

PT otitis media, bacteraemia and/or pneumonia -

XX Example 13; Fig 45; 106pp; English.

XX The present invention describes nucleic acids (I) encoding protein

CC antigens (II) from Streptococcus pneumoniae. The protein antigens

CC have bactericidal activity. The nucleic acids, encoding the proteins

CC antigens, may be used for the recombinant production of the proteins

CC they encode. The protein antigens may then be used as vaccines for the

CC prevention and treatment of Streptococcal infections in mammals

CC (especially humans) which result in, e.g. meningitis, otitis media,

CC bacteraemia and/or pneumonia. The present sequence encodes the

CC S. pneumoniae GBS BVH-71 protein antigen.

XX Sequence 2469 BP; 842 A; 481 C; 519 G; 627 T; 0 other;

SO Query Match 0.8%; Score 20; DB 21; Length 2469;

Best Local Similarity 100.0%; Pred. No. 23;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 479 CAAGACGCTACTACTACAGA 498

DB 571 CAAGACGCTACTACTACAGA 590

XX RESULT 63

AAA05811

ID AAA05811 standard; DNA; 2469 BP.

XX AAA05811;

AC 30-MAY-2000 (first entry)

DT Group B Streptococcus protein encoding nucleotide sequence SEQ ID NO:9.

XX Group B Streptococcus; Streptococcus agalactiae; protein antigen;

DE vaccine; screening; immunogen; detection; diagnosis; infection;

XX antibody; antibody; antibacterial; ds.

KW Streptococcus agalactiae.

OS

```

PN WO200006736-A2.
XX
XX
PD 10-FEB-2000.
XX
XX
PF 27-JUL-1999; 99WO-GB02444.
XX
XX
PR 27-JUL-1998; 98GB-0016335.
PR 19-MAR-1999; 99US-0125163.
XX
XX
PA (MICR-) MICROBIAL TECHNIQS LTD.
XX
XX
PI Le Page RW, Wells JM, Hanniffy SB;
XX
XX
DR WPI; 2000-195299/17.
DR P-PSDB; AA91283.
XX
XX
PT New Group B Streptococcus protein, useful as vaccine, for diagnosis of
PT Streptococcal infections and for screening of antibodies or affibodies
XX
XX
PS Claim 4; Fig 1; 123pp; English.
XX
XX
CC AAA05803 to AAA05872 encode proteins, polypeptides and peptides (given
CC in AA91275 to AA91343) isolated from Group B Streptococcus (GBS), also
CC known as Streptococcus agalactiae. The GBS polynucleotides and
CC polypeptides have antibacterial activity. Immunogenic compositions
CC comprising GBS polynucleotides or polypeptides can be used as vaccines
CC and for the treatment or prophylaxis of GBS infection. The
CC polynucleotides and polypeptides can also be used in the detection of GBS
CC and for screening DNA encoding bacterial cell envelope associated or
CC secreted antigens in gram positive bacteria. AA05873 to AA05941
CC represent primers used in the exemplification of the present invention.
SQ
SQ Sequence 2469 BP; 842 A; 477 C; 514 G; 636 T; 0 other;

Query Match 0.8%; Score 20; DB 21; Length 2469;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 479 CAAGGACGCTATCTACAGA 498
Db 571 CAAGGACGCTATCTACAGA 590

RESULT 64
AAS00038
ID AAS00038 standard; DNA; 2469 BP.
XX
XX
AC AAS00038;
XX
XX
DT 11-MAY-2001 (first entry)
XX
XX
DE Streptococcus agalactiae DNA encoding GBS36.
XX
XX
KW Group B streptococci protein 36; GBS36; immunogen; vaccine;
KW antibody; necrotising fasciitis; scarlet fever; sepsis; impetigo;
KW bacterial meningitis; otitis media; community-acquired pneumonia; ds.
XX
XX
OS Streptococcus agalactiae.
XX
XX
FH Key Location/Qualifiers
FT CDS 1..2469
FT /tag= a
FT /product= "GBS36"
XX
XX
PN WO200114421-A1.
XX
XX
PD 01-MAR-2001.
XX
XX
PF 25-AUG-2000; 2000WO-US23417.
XX
XX
PR 25-AUG-1999; 99US-0150750.
XX
XX
PA (MED-) MEDIMUNE INC.

```

```

XX
XX
PI Koenig S, Heinrichs J, Johnson LS, Adamou JE;
XX
XX
DR WPI; 2001-211305/21
DR P-PSDB; AAU00028, AAU00030.
XX
XX
XX
XX
PT New polypeptides obtained from group A or B streptococci, especially
PT Streptococcus aureus homologous to Sp36 protein of Streptococcus
PT pneumoniae useful as antibacterial vaccines
XX
XX
PS Claim 10; Page 53-54; 62pp; English.
XX
XX
CC The sequence encodes Streptococcus agalactiae Group B Streptococci
CC protein 36, GBS36. A recombinant cell producing GBS36, GBS36(2) or
CC GBS36 is useful as a vaccine for vaccinating an animal, preferably a
CC human against infection by a bacterial organism such as a streptococcal
CC or staphylococcal bacteria, and for treating a disease caused by group A
CC streptococci, group B streptococci or Staphylococcus aureus in an animal
CC preferably a human. Vaccines and antibodies against the proteins of the
CC invention are useful in prophylaxis and/or treatment of diseases such as
CC necrotising fasciitis, scarlet fever, sepsis, impetigo, bacterial
CC meningitis, otitis media, community-acquired pneumonia and many diseases
CC of newborns. The proteins are also used as immunogens to stimulate the
CC production of antibodies for use in passive immunotherapy, for use as
CC diagnostic reagents and for use as reagents in other processes such as
CC affinity chromatography.
XX
XX
SQ Sequence 2469 BP; 844 A; 477 C; 514 G; 634 T; 0 other;

Query Match 0.8%; Score 20; DB 22; Length 2469;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 479 CAAGGACGCTATCTACAGA 498
Db 571 CAAGGACGCTATCTACAGA 590

RESULT 65
AAA65741
ID AAA65741 standard; DNA; 2472 BP.
XX
XX
AC AAA65741;
XX
XX
DT 21-NOV-2000 (first entry)
XX
XX
DE Streptococcus pneumoniae GAS BVH-71 gene SEQ ID NO:82.
XX
XX
XX
XX
KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal; ds.
XX
XX
OS Streptococcus pneumoniae.
XX
XX
PN WO2000039299-A2.
XX
XX
PD 06-JUL-2000.
XX
XX
PF 20-DEC-1999; 99WO-CA01218.
XX
XX
PR 23-DEC-1998; 98US-0113800.
XX
XX
PA (BIOC-) BIOCHEM PHARMA INC.
XX
XX
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX
XX
DR WPI; 2000-452397/39.
DR P-PSDB; AAB12747.
XX
XX
PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia
XX
XX
PS Example 13; Fig 47; 106pp; English.

```

XX The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence encodes the
CC S. pneumoniae GAS BVH-71 protein antigen.
XX
SQ Sequence 2472 BP, 846 A, 480 C, 514 G, 632 T, 0 other;
Query Match 0.8%; Score 20; DB 21; Length 2472;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 479 CAAGAGCGCTACTACTACAGA 498
Db 571 CAAGAGCGCTACTACTACAGA 590
|||||
RESULT 66
ABN66838 standard; DNA, 2475 BP.
XX
AC ABN66838;
XX
DT 01-JUL-2002 (first entry)
XX
DE Streptococcus polynucleotide SEQ ID NO 1589.
XX
XX Streptococcus GAS; GAS; group B streptococcus; Streptococcus agalactiae;
KM group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KM antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
OS Streptococcus pyogenes.
XX
PN WO200234771-A2.
XX
PD 02-MAY-2002.
XX
PF 29-OCT-2001; 2001WO-GB04789.
XX
PR 27-OCT-2000; 2000GB-0026333.
PR 24-NOV-2000; 2000GB-0028727.
PR 07-MAR-2001; 2001GB-0005640.
XX
XX (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
PI Tettelin H;
XX
DR WPI: 2002-352536/38.
DR P-PSDB; ABP26207.
XX
XX New Streptococcus protein for the treatment or prevention of infection
PT or disease caused by Streptococcus bacteria, such as meningitis, and
PT for detecting a compound that binds to the protein -
XX
PS Claim 7; Page 3309; 4525pp; English.
XX
CC The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to

CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.
XX
SQ Sequence 2475 BP, 844 A, 486 C, 518 G, 627 T, 0 other;
Query Match 0.8%; Score 20; DB 24; Length 2475;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 479 CAAGAGCGCTACTACTACAGA 498
Db 571 CAAGAGCGCTACTACTACAGA 590
|||||
RESULT 67
AAS00036 standard; DNA, 2478 BP.
XX
ID AAS00036;
XX
AC AAS00036;
XX
DT 11-MAY-2001 (first entry)
XX
DE Streptococcus pyogenes DNA encoding GAS36.
XX
XX Group A streptococci protein 36; GAS36; immunogen; vaccine;
XX antibody; necrotising fasciitis; scarlet fever; sepsis; impetigo;
KM bacterial meningitis; otitis media; community-acquired pneumonia; ds.
XX
OS Streptococcus pyogenes.
XX
XX
XX Key Location/Qualifiers
FH 1..2478
FT CDS /*tag= a
FT /product= "GAS36"
XX
XX WO200114421-A1.
XX
PN 01-MAR-2001.
XX
PD 25-AUG-2000; 2000WO-US23417.
XX
PE 25-AUG-1999; 99US-0150750.
XX
PR (MEDI-) MEDIMUNE INC.
XX
PA Koenig S, Heinrichs J, Johnson LS, Adamou JE;
PI
XX WPI: 2001-211305/21.
DR P-PSDB; AAU00026, AAU00029.
XX
XX New polypeptides obtained from group A or B streptococci, especially
PT Streptococcus aureus homologous to Sp36 protein of Streptococcus
PT pneumoniae useful as antibacterial vaccines -
XX
XX Claim 10; Page 46-47; 62pp; English.
XX
CC The sequence encodes Streptococcus pyogenes Group A Streptococci
CC protein 36, GAS36. A recombinant cell producing GAS36, GAS36(2) or
CC GAS36 is useful as a vaccine for vaccinating an animal, preferably a
CC human against infection by a bacterial organism such as a streptococcal
CC or staphylococcal bacteria, and for treating a disease caused by group A
CC streptococci, group B streptococci or staphylococcus aureus in an animal
CC preferably a human. Vaccines and antibodies against the proteins of the
CC invention are useful in prophylaxis and/or treatment of diseases such as
CC necrotising fasciitis, scarlet fever, sepsis, impetigo, bacterial
CC meningitis, otitis media, community-acquired pneumonia and many diseases
CC of newborns. The proteins are also used as immunogens to stimulate the
CC production of antibodies for use in passive immunotherapy, for use as

CC diagnostic reagents and for use as reagents in other processes such as
CC affinity chromatography.

XX Sequence 2478 BP; 846 A; 486 C; 518 G; 628 T; 0 other;

XX Query Match 0.8%; Score 20; DB 22; Length 2478;
XX Best Local Similarity 100.0%; Pred. No. 23;
XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 479 CAAGACGCTATCTACAGA 498
DB 571 CAAGACGCTATCTACAGA 590

RESULT 68
AA169348/c
ID AA169348 standard; DNA; 5158 BP.

XX AA169348;

XX 18-FEB-2002 (first entry)

DE C. albicans contig4-2069 DNA containing hyphal-specific proteins.

XX Hyphal-specific protein; biochip; diagnosis; infection; Cap18p; Cap19p;
XX screening; fungicide; virulent; ds.

XX Candida albicans.

OS WO200185989-A2.

XX 15-NOV-2001.

XX 10-MAY-2001; 2001WO-EP05363.

XX 11-MAY-2000; 2000DE-1023130.

XX (FRAU) FRAUNHOFER GBS FOERDERUNG ANGEWANDTEN.

XX Rupp S, Johannes F, Sohn K;

XX WPI; 2002-062252/08.

XX Nucleotide chip for specific detection of the hyphal, virulent, form of
XX Candida albicans, comprises a solid support carrying a nucleotide
XX sequence for identifying and transcribing genes that encode
XX hyphae-specific proteins

XX Disclosure; Page 60-61; 65pp; German.

XX This invention describes a novel nucleotide chip which comprises a
XX solid support carrying a nucleotide sequence for identifying and
XX transcribing genes that encode hyphal-specific proteins. The chip of the
XX invention, and related protein and antibody chips, are used: (i) for
XX diagnosis of Candida-mediated diseases (local or systemic infections);
XX and (ii) to screen for agents that inhibit expression of hyphae-specific
XX proteins, potential therapeutic agents. The products of the invention
XX have fungicidal activity, (i), or related proteins and antibodies, are
XX specific for the hyphal (virulent) form of Candida albicans, so allow
XX differentiation between virulent and avirulent strains. This sequence
XX represents the Candida albicans derived contig4-2069 DNA which encodes
XX the Cap18p and Cap19p hyphal-specific protein described in the method
XX of the invention.

XX Sequence 5158 BP; 1762 A; 768 C; 781 G; 1847 T; 0 other;

XX Query Match 0.8%; Score 20; DB 24; Length 5158;
XX Best Local Similarity 100.0%; Pred. No. 23;
XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1512 CAATTTCTCAATTCAGTAT 1531
DB 4215 CAATTTCTCAATTCAGTAT 4196

RESULT 69

XX AAX91105 standard; DNA; 5215 BP.

XX AAX91105;

XX 15-NOV-1999 (first entry)

DE Group B Streptococcus (GBS) antigens encoding DNA (clone 3).

XX Group B Streptococcus; GBS; antigen; vaccine; Streptococcus infection;
XX sepsis; meningitis; pneumonia; immunocompromise; diabetes; liver disease;
XX cancer; veterinary; mastitis; ss.

OS Streptococcus sp.

XX Key Location/Qualifiers

XX CDS 3..125

XX /tag= a

XX /note= "see AAY27346"

XX 133..2514

XX /tag= b

XX /note= "see AAY27347"

XX 367..2514

XX /tag= c

XX /note= "see AAY27348"

XX complement (2716..2946)

XX /tag= d

XX /note= "see AAY27349"

XX complement (2995..3252)

XX /tag= e

XX /note= "see AAY27350"

XX complement (3299..3676)

XX /tag= f

XX /note= "see AAY27351"

XX complement (3837..4124)

XX /tag= g

XX /note= "see AAY27352"

XX complement (4351..5214)

XX /tag= h

XX /note= "see AAY27353"

XX MO9942588-A2.

XX 26-AUG-1999.

XX 17-FEB-1999; 99WO-CA00114.

XX 20-FEB-1999; 98US-0075425.

XX (BIOC-) BIOCHEM VACCINS INC.

XX Boyer M, Brodeur BR, Charlebois I, Hamel J, Martin D;

XX Rieux C;

XX WPI; 1999-540309/45.

XX P-PDB; AAY27346, AAY27347, AAY27348, AAY27349, AAY27350, AAY27351,
XX AAY27352, AAY27353.

XX Novel group B Streptococcus antigens - useful as vaccine
XX compositions for prophylaxis or therapy of Streptococcus infections

XX Claim 10; Fig 3A; 154pp; English.

XX The invention provides Group B Streptococcus (GBS) antigens
XX (AAY27336-370) and nucleic acids (AAX91103-X91111) encoding the antigens.
XX The GBS antigens can be recombinantly expressed using standard
XX recombinant methodology. The GBS antigens of the invention can be used as
XX vaccine components for the treatment or prophylaxis of diseases and
XX symptoms mediated by Streptococcus infection, especially group A
XX Streptococcus (S. pyogenes), GBS or S. agalactiae, S. dysgalactiae, S.

CC uberis, *S. nodocardia*, as well as *Staphylococcus aureus*. The vaccines are
CC administered to those individuals at risk of GBS infection, particularly
CC pregnant women and infants for sepsis, meningitis, and pneumonia, as well
CC as immunocompromised individuals, such as those with diabetes, liver
CC disease or cancer. The vaccines also have veterinary applications, such as
CC for the treatment of mastitis in cattle. The present sequence represents
CC a DNA encoding 8 GBS antigens of the invention.
XX
SQ Sequence 5215 BP; 1631 A; 957 C; 1038 G; 1589 T; 0 other;
Query Match 0.8%; Score 20; DB 20; Length 5215;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 479 CAAGGACGCTATCTACAGCA 498
DB 616 CAAGGACGCTATCTACAGCA 635
RESULT 70
ID ABX54820 standard; cDNA; 164 BP.
XX
AC ABX54820;
XX
DT 26-FEB-2003 (first entry)
XX
DE Bovine EST associated with lactation/muscle/fat deposition #4749.
XX
KW Bovine; ss; EST; expressed sequence tag; lactation; LMPD;
KW muscle deposition; fat deposition; genome mapping; gene identification;
KW gene analysis; cattle breeding.
XX
OS Bos Taurus.
XX
PN US2002137160-A1.
XX
PD 26-SEP-2002.
XX
PE 26-OCT-2001; 2001US-0983965.
XX
PR 17-DEC-1998; 98US-113678P.
PR 15-DEC-1999; 99US-0465231.
XX
PA (BYAT/) BYATT J C.
PA (MATH/) MATHIALAGAN N.
PA (TAON/) TAO N.
PA (WARR/) WARREN W C.
PI Byatt JC, Mathialagan N, Tao N, Warren WC;
XX
DR WPI; 2003-102386/09.
XX
PT Purified nucleic acid molecules, useful for genome mapping, gene
PT identification and analysis, cattle breeding or preparation of
PT constructs for cattle gene expression and genetically improved cattle -
XX
PS Claim 2; SEQ ID No 4749; 38pp; English.
XX
CC The invention relates to a purified nucleic acid molecule associated with
CC lactation or muscle and fat deposition (designated LMPD), derived
CC from cattle, and the LMPD nucleic acid can specifically hybridise to a
CC second nucleic acid molecule comprising any of 5912 nucleotide
CC sequences, appearing as ABX50072-ABX55983, or complements of them.
CC Also included are: (1) a transformed cell having a nucleic acid
CC comprising an LMPD nucleic acid linked to a promoter and a 3' non-
CC translated sequence that functions in the cell to cause termination of
CC transcription and addition of polyadenylated ribonucleotides to a 3' end
CC of the mRNA molecule; and (2) determining a level or pattern of a
CC molecule in a bovine cell or tissue comprising: (a) incubating a marker
CC nucleic acid (comprising any of the 5912 nucleic acid sequences or its
CC complement or fragment) with a complementary nucleic acid molecule
CC obtained from the bovine cell or tissue, where hybridisation between the

CC marker nucleic acid and the complementary nucleic acid permits the
CC detection of the molecule; and (b) detecting the level or pattern of the
CC complementary nucleic acid, where the detection of the complementary
CC nucleic acid is predictive of the level or pattern of the molecule.
CC The LMPD nucleic acid is used for determining a level or pattern
CC of a molecule in a bovine cell or tissue. It is useful for genome
CC mapping, gene identification and analysis, cattle breeding, preparation
CC of constructs for use in cattle gene expression, or for genetically
CC improving cattle. The present sequence is one of the 5912 bovine
CC LMPD EST (expressed sequence tag) nucleic acids.
CC Note: The present sequence was not shown in the specification but
CC was obtained in electronic format from the USPTO web site:
CC seqdata.uspto.gov/sequence.html?docid=20020137160.
XX
SQ Sequence 164 BP; 60 A; 41 C; 34 G; 29 T; 0 other;
Query Match 0.8%; Score 19; DB 25; Length 164;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1112 TCAATCTCTTGTGTTA 1130
DB 83 TCAATCTCTTGTGTTA 65
RESULT 71
ID ABV02215 standard; cDNA; 426 BP.
XX
AC ABV02215;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 2206.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PE 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 432; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer

CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for:
 CC (a) assessing whether a patient is afflicted with prostate cancer;
 CC (b) monitoring the progression of prostate cancer in a patient;
 CC (c) assessing the efficacy of a test compound to inhibit prostate
 CC cancer in a patient;
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
 CC in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound;
 CC (g) determining whether prostate cancer has metastasized in a patient;
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a
 CC patient;
 CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
 CC
 SO Sequence 468 BP; 95 A; 78 C; 74 G; 220 T; 1 other;
 Query Match 0.8%; Score 19; DB 23; Length 468;
 Best Local Similarity 100.0%; Pred. No. 68;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1224 TTTAAATCTGAAAGCAA 1242
 Db 136 TTTAAATCTGAAAGCAA 118
 RESULT 74
 ABV32529/C
 ID ABV32529 standard; cDNA; 481 BP.
 XX
 AC ABV32529;
 XX
 DT 16-SEP-2002 (first entry)
 XX
 DE Human prostate expression marker cDNA 32520.
 XX
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200160860-A2.
 XX
 PD 23-AUG-2001.
 XX
 PF 20-FEB-2001; 2001WO-US05171.
 XX
 PR 17-FEB-2000; 2000US-183319P.
 PR 16-MAR-2000; 2000US-189862P.
 PR 25-MAY-2000; 2000US-207454P.
 PR 09-JUN-2000; 2000US-211314P.
 PR 18-JUL-2000; 2000US-219007P.
 PR 13-DEC-2000; 2000US-255281P.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Schlegel R, Endege WO, Monahan JE;
 XX
 DR WPI; 2001-662795/76.
 XX
 PT Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer -
 XX
 PS Claim 1; Page 6937; 11750pp; English.
 XX
 CC The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for:
 CC (a) assessing whether a patient is afflicted with prostate cancer;
 CC (b) monitoring the progression of prostate cancer in a patient;
 CC (c) assessing the efficacy of a test compound to inhibit prostate
 CC cancer in a patient;

CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
 CC in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound;
 CC (g) determining whether prostate cancer has metastasized in a patient;
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a
 CC patient;
 CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
 CC
 SO Sequence 481 BP; 101 A; 81 C; 85 G; 213 T; 1 other;
 Query Match 0.8%; Score 19; DB 23; Length 481;
 Best Local Similarity 100.0%; Pred. No. 68;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1224 TTTAAATCTGAAAGCAA 1242
 Db 181 TTTAAATCTGAAAGCAA 163
 RESULT 75
 ABV41454/C
 ID ABV41454 standard; cDNA; 481 BP.
 XX
 AC ABV41454;
 XX
 DT 16-SEP-2002 (first entry)
 XX
 DE Human prostate expression marker cDNA 41445.
 XX
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200160860-A2.
 XX
 PD 23-AUG-2001.
 XX
 PF 20-FEB-2001; 2001WO-US05171.
 XX
 PR 17-FEB-2000; 2000US-183319P.
 PR 16-MAR-2000; 2000US-189862P.
 PR 25-MAY-2000; 2000US-207454P.
 PR 09-JUN-2000; 2000US-211314P.
 PR 18-JUL-2000; 2000US-219007P.
 PR 13-DEC-2000; 2000US-255281P.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Schlegel R, Endege WO, Monahan JE;
 XX
 DR WPI; 2001-662795/76.
 XX
 PT Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer -
 XX
 PS Claim 1; Page 8330; 11750pp; English.
 XX
 CC The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for:
 CC (a) assessing whether a patient is afflicted with prostate cancer;
 CC (b) monitoring the progression of prostate cancer in a patient;
 CC (c) assessing the efficacy of a test compound to inhibit prostate
 CC cancer in a patient;
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
 CC in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound;
 CC (g) determining whether prostate cancer has metastasized in a patient;
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a

CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ Sequence 481 BP; 101 A; 81 C; 85 G; 213 T; 1 other;

Query Match 0.8%; Score 19; DB 23; Length 481;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1224 TTTAAATCTTGAAGCAA 1242
DB 181 TTTAAATCTTGAAGCAA 163

RESULT 76
AB089355/C
ID AB089355 standard; cDNA; 693 BP.

XX AB089355;

DT 27-SEP-2002 (first entry)

DE Human prostate expressed polynucleotide SEQ ID NO 611.

KW Human; prostate; cytostatic; tumour; cancer; vaccine; gene therapy;

KW gene; ss.

OS Homo sapiens.

FN WO200255700-A2.

PD 18-JUL-2002.

PF 07-DEC-2001; 2001WO-US47349.

PR 07-DEC-2000; 2000US-254648P.

PR 13-MAR-2001; 2001US-275688P.

PA (CHIR) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

PI Escobedo J, Garcia PD, Kassam A, Lamson G, Drmanac R;

PI Ckrnenjakov R, Dickson M, Drmanac S, Labat I, Leshkowitz D;

PI Kita D, Garcia V, Jones WL, Stache-Grain B, Scott EM;

DR WPI; 2002-557824/59.

PT New genes and gene products isolated from human prostate, useful for
PT treating or diagnosing tumor or cancer (e.g. prostate cancer or breast
PT cancer), or as vaccines for treating or preventing these diseases -

PS Claim 1; SEQ ID NO 611; 186pp + Sequence Listing; English.

CC The invention relates to an isolated polynucleotide comprising any of
CC 1477 sequences or its fragment, degenerate variant, antisense or
CC complement. The polynucleotides and gene products are useful for treating
CC or diagnosing tumor or cancer (e.g. prostate cancer, breast cancer,
CC lung cancer or medullary carcinoma) in a subject (e.g. cattle, dogs,
CC cats, rabbits, horse or human). The polynucleotides and polypeptides are
CC also useful as vaccines for treating or preventing these diseases. The
CC polynucleotides are useful for gene therapy. The present sequence is that
CC of one of a group of polynucleotides (AB088745-AB090015) disclosed
CC electronically as sequences of the invention. However only 1271
CC polynucleotide sequences are given, whereas 1477 polynucleotides and 91
CC proteins are claimed.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequence.

XX Sequence 693 BP; 239 A; 88 C; 115 G; 235 T; 16 other;

Query Match 0.8%; Score 19; DB 24; Length 693;
Best Local Similarity 100.0%; Pred. No. 69;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1734 TCTTACCTCCATCTCCAGA 1752
DB 155 TCTTACCTCCATCTCCAGA 137

RESULT 77
ABV21401
ID ABV21401 standard; cDNA; 1203 BP.

XX ABV21401;

DT 13-SEP-2002 (first entry)

DE Human prostate expression marker cDNA 21392.

KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;

KW pharmacogenomic marker; gene; ss.

OS Homo sapiens.

FN WO200160860-A2.

PD 23-AUG-2001.

PF 20-FEB-2001; 2001WO-US05171.

PR 17-FEB-2000; 2000US-18319P.

PR 16-MAR-2000; 2000US-189862P.

PR 25-MAY-2000; 2000US-207454P.

PR 09-JUN-2000; 2000US-211314P.

PR 18-JUL-2000; 2000US-219007P.

PR 13-DEC-2000; 2000US-255281P.

PA (MILL-) MILENNIUM PREDICTIVE MEDICINE INC.

PI Schlegel R, Endege WO, Monahan JE;

DR WPI; 2001-662795/76.

PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -

PS Claim 1; Page 3561; 11750pp; English.

CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:

CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;

CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;

CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;

CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX Sequence 1203 BP; 545 A; 170 C; 185 G; 302 T; 1 other;

Query Match 0.8%; Score 19; DB 23; Length 1203;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1224 TTTAAATCTTGAAGCAA 1242
DB 557 TTTAAATCTTGAAGCAA 575

RESULT 78
 ABV27219
 ID ABV27219 standard; cDNA; 1205 BP.
 XX
 AC ABV27219;
 XX
 DT 16-SEP-2002 (first entry)
 XX
 DE Human prostate expression marker CDNA 27210.
 XX
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 XX
 OS Homo sapiens.
 XX
 FN WO200160860-A2.
 XX
 PD 23-AUG-2001.
 XX
 PE 20-FEB-2001; 2001WO-US05171.
 XX
 PR 17-FEB-2000; 2000US-18319P.
 PR 16-MAR-2000; 2000US-189662P.
 PR 25-MAY-2000; 2000US-207454P.
 PR 09-JUN-2000; 2000US-211314P.
 PR 18-JUL-2000; 2000US-21907P.
 PR 13-DEC-2000; 2000US-255281P.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Schlegel R, Endege WO, Monahan JE;
 XX
 WPI; 2001-662795/76.
 XX
 DR Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer -
 XX
 XX Claim 1; Page 5523; 11750P; English.
 XX
 CC The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for:
 CC (a) assessing whether a patient is afflicted with prostate cancer;
 CC (b) monitoring the progression of prostate cancer in a patient;
 CC (c) assessing the efficacy of a test compound to inhibit prostate
 CC cancer in a patient;
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
 CC in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound;
 CC (g) determining whether prostate cancer has metastasized in a patient;
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a
 CC patient;
 CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
 XX
 SO Sequence 1205 BP; 547 A; 170 C; 185 G; 302 T; 1 other;
 XX
 Query Match 0.8%; Score 19; DB 23; Length 1205;
 Best Local Similarity 100.0%; Pred. No. 71;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1224 TTTAAATCTTGAAAGCAA 1242
 DB 559 TTTAAATCTTGAAAGCAA 577
 XX
 RESULT 79
 AAA05838
 ID AAA05838 standard; DNA; 1455 BP.
 XX
 AC AAA05838;

XX
 DT 30-MAY-2000 (first entry)
 XX
 DE Group B Streptococcus protein encoding nucleotide³ sequence SEQ ID NO:38.
 XX
 KW Group B Streptococcus; Streptococcus agalactiae; protein antigen;
 KW vaccine; screening; immunogen; detection; diagnosis; infection;
 KW antibody; affibody; antibacterial; ds.
 XX
 OS Streptococcus agalactiae.
 XX
 FN WO200006736-A2.
 XX
 PD 10-FEB-2000.
 XX
 PE 27-JUL-1999; 99WO-GB02444.
 XX
 PR 27-JUL-1998; 98GB-0016335.
 PR 19-MAR-1999; 99US-0125163.
 XX
 PA (MICR-) MICROBIAL TECHNIQS LTD.
 XX
 PI Le Page RWF, Wells JM, Hamifly SB;
 XX
 WPI; 2000-195299/17.
 DR P-PSDB; AAY91310.
 XX
 PT New Group B Streptococcus protein, useful as vaccine, for diagnosis of
 PT Streptococcal infections and for screening of antibodies or affibodies
 XX
 PS Claim 4; Fig 1; 123P; English.
 XX
 CC AAA05803 to AAA05872 encode proteins, polypeptides and peptides (given
 CC in AAY91275 to AAY91343) isolated from Group B streptococcus (GBS), also
 CC known as Streptococcus agalactiae. The GBS polynucleotides and
 CC polypeptides have antibacterial activity. Immunogenic compositions
 CC comprising GBS polynucleotides or polypeptides can be used as vaccines
 CC and for the treatment or prophylaxis of GBS infection. The
 CC polynucleotides and polypeptides can also be used in the detection of GBS
 CC and for screening DNA encoding bacterial cell envelope associated or
 CC secreted antigens in gram positive bacteria. AAA05873 to AAA05941
 CC represent primers used in the exemplification of the present invention.
 XX
 SO Sequence 1455 BP; 584 A; 237 C; 271 G; 363 T; 0 other;
 XX
 Query Match 0.8%; Score 19; DB 21; Length 1455;
 Best Local Similarity 100.0%; Pred. No. 71;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1583 GATGTTACATTTTGGATG 1601
 DB 343 GATGTTACATTTTGGATG 361
 XX
 RESULT 80
 AAA70231/c
 ID AAA70231 standard; DNA; 1650 BP.
 XX
 AC AAA70231;
 XX
 DT 07-NOV-2000 (first entry)
 XX
 DE Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:364.
 XX
 KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
 KW antimalarial; malaria; protozoacide; infection; insecticide; ds.
 XX
 OS Plasmodium falciparum.
 XX
 FN WO200025728-A2.
 XX
 PD 11-MAY-2000.
 XX

```

PF 05-NOV-1999; 99WO-US26796.
XX
XX 05-NOV-1998; 98US-0107131.
XX
XX (HOPE/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
XX Hoffman S, Carucci D, Gardner M, Venter JC;
XX WPI; 2000-365347/31.
XX
XX Proteins encoded by chromosome 2 of the human malarial parasite,
XX Plasmodium falciparum, useful as antimalarial vaccines and in the
XX diagnosis of P.falciparum infection -
XX
XX Disclosure; Page 551; 577pp; English.
XX
XX The present invention describes proteins and their fragments (I) encoded
XX by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
XX Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
XX vaccines against P. falciparum infection comprising (I) or (II).
XX (I) and (II) are useful for the development of vaccines against
XX P. falciparum infection. (I) and polyclonal antisera or a monoclonal
XX antibody raised to immunogens comprising the sequences of (I), are
XX useful in the detection of infection with P. falciparum. Furthermore,
XX (I) (especially when they are refined or secreted or membrane proteins)
XX can aid the identification of drugs to treat or prevent P. falciparum
XX infection, or they can be used to identify drug resistance in
XX P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
XX subsequent identification of proteins encoded by it will help to expand
XX our understanding of parasite biology, a process hampered by the
XX complexity of the parasitic lifecycle, and provide new targets for
XX vaccine and drug development. Parasite resistance to drugs and mosquito
XX resistance to insecticides have led to a resurgence of malaria in many
XX parts of the world, and there is a pressing need for vaccines and new
XX drugs. AAV70078 to AAV70287 and AAB18144 to AAB18352 represent nucleotide
XX and protein sequences given in the present invention, but which are not
XX specifically mentioned within the specification.
XX
XX Sequence 1650 BP, 819 A; 147 C; 245 G; 439 T; 0 other;
SQ
Query Match 0.8%; Score 19; DB 21; Length 1650;
Best Local Similarity 100.0%; Pred. NO. 72;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 505 TTATATCTTAAATGCTTCT 523
Db 121 TTATATCTTAAATGCTTCT 103

```

```

PR 20-MAY-1999; 99US-0315179.
XX
XX (UYEM-) UNIV EMORY.
XX
XX Lollar JS;
XX
XX WPI; 2001-016350/02.
XX P-PSDB; AAB50467.
XX
XX A modified human factor VIII useful for treating hemophiliacs with
XX immunity to human factor VIII comprises insertion of immunoreactivity
XX reducing amino acid into the factor VIII sequence -
XX
XX Disclosure; Page 120-124; 172pp; English.
XX
XX The present sequence is given in a specification relating to a modified
XX human factor VIII (FVIII) comprising a 2332 amino acid sequence
XX containing one or more substitutions at positions 484-508 which are
XX insertions of an immunoreactivity reducing amino acid. The modified
XX factor VIII has procoagulant activity. The protein is useful for treating
XX patients with a factor VIII deficiency who have developed antibodies that
XX inhibit the activity of factor VIII. The modified factor VIII molecule
XX evades immune detection in such patients.
XX
XX Sequence 7492 BP, 2487 A; 1503 C; 1436 G; 2066 T; 0 other;
SQ
Query Match 0.8%; Score 19; DB 22; Length 7492;
Best Local Similarity 100.0%; Pred. NO. 76;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2347 AAAGAGATATCTTCA 2365
Db 206 AAAGAGATATCTTCA 224

```

```

RESULT 81
AAC90510
ID AAC90510 standard; cDNA; 7492 BP.
XX
XX AAC90510;
XX
XX 15-MAR-2001 (first entry)
XX
XX Mouse factor VIII cDNA.
XX
XX Mouse; factor VIII; cDNA; coagulant; clotting factor; haemophilic;
XX immune detection evasion; ss.
XX
XX Mus musculus.
XX
XX WO200071141-A1.
XX
XX 30-NOV-2000.
XX
XX 16-MAY-2000; 2000WO-US13541.
XX
XX

```

```

RESULT 82
AAV25812
ID AAV25812 standard; cDNA; 7493 BP.
XX
XX AAV25812;
XX
XX 10-JUL-1998 (first entry)
XX
XX Murine factor VIII encoding cDNA.
XX
XX Murine; factor VIII; hybrid; haemophilic; procoagulant; blood; clot; ss.
XX
XX Mus musculus.
XX
XX Key Location/Qualifiers
XX repeat_unit 1..407
XX /tag= a
XX /rpt_type= TERMINAL
XX /note= "5'UTR"
XX CDS 408..7367
XX /tag= b
XX /product= "coagulation factor VIII"
XX repeat_unit 7368..7476
XX /tag= c
XX /rpt_type= TERMINAL
XX /note= "3'UTR"
XX polyA_signal 7471..7476
XX /tag= d
XX
XX US5744446-A.
XX
XX 28-APR-1998.
XX
XX 07-JUN-1995; 95US-0474503.
XX
XX 07-JUN-1995; 95US-0474503.
XX
XX 07-APR-1992; 92US-0864004.
XX
XX 11-MAR-1994; 94US-0212133.
XX

```

PR 15-NOV-1994; 94WO-US13200.
XX (UYEM-) UNIV EMORY.
XX Loliar JS, Runge MS;
XX WPI; 1998-271107/24.
DR P-PSDB; AAW53485.
XX
XX Hybrid of human and animal factor VIII - containing porcine and
PT murine amino acid sequences is useful in the treatment of
PT haemophilia
XX
XX Disclosure; Column 65-72; 48pp; English.
XX
XX The present sequence encodes murine factor VIII used in the present
CC invention. The present invention describes a new procoagulant hybrid
CC factor VIII (I) comprising human factor VIII and has amino acid
CC sequences substituted from the group of A2 domain fragments consisting
CC of amino acids 373-540, 373-508, 445-508, 484-508, 489-508
CC and 484-489 from the human 2332 amino acid A2 domain sequence (II) as
CC given in the specification (see AAW53483). The substitution is from
CC corresponding non-human mammalian factor VIII sequences. Also described
CC is a method for treating factor VIII deficiency comprising administering
CC a therapeutically effective dose of (I) in a pharmaceutical carrier.
CC (II), prepared from reconstitution of purified molecules or recombinant
CC techniques, is useful in the treatment of haemophiliacs who have factor
CC VIII deficiencies and whose blood is not normally able to clot after
CC internal or external bleeding. (I) compared to native human factor VIII
CC is more stable at physiological conditions and has a higher specific
CC clotting activity.
XX
SQ Sequence 7493 BP; 2487 A; 1503 C; 1436 G; 2067 T; 0 other;
Query Match 0.8%; Score 19; DB 19; Length 7493;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2347 AAAGAGTAATCCTTCA 2365
Db 206 AAAGAGTAATCCTTCA 224
RESULT 83
AAV12115
ID AAV12115 standard; cDNA to mRNA; 7493 BP.
XX
AC AAV12115;
XX
DT 17-JUN-1998 (first entry)
XX
DE Mus musculus factor VIII coding region.
XX
XX factor VIII; recombinant; modified; haemophilia; treatment; ds.
XX
XX Mus musculus.
XX
XX Key Location/Qualifiers
FH 1..407
FT 5'UTR
FT /*tag= a
FT 7368..7493
FT /*tag= b
FT polyA_signal 7471..7476
FT /*tag= c
FT CDS 408..7367
FT /*tag= d
PT /product= factor VIII

XX
XX 26-JUN-1996; 96US-0670707.
PR
XX (UYEM-) UNIV EMORY.
XX
XX Loliar JS;
XX WPI; 1998-077108/07.
DR P-PSDB; AAW44135.
XX
XX New modified factor VIII molecules - having reducing immunogenicity
PT
XX
XX Disclosure; Pages 74-77; 126pp; English.
XX
XX The sequence is that encoding murine factor VIII. It can be used in
CC the production of modified factor VIII. The factor VIII molecules
CC have coagulant activity and can be used for treating factor VIII
CC deficiency, particularly for treating patients with haemophilia.
CC The products can also be used in detection and diagnosis. This
CC modified factor VIII has less immunoreactivity with naturally
CC occurring inhibitory antibodies to factor VIII and may be less apt
CC to elicit the production of antibodies to factor VIII than human
CC factor VIII. Some of the hybrid factor VIII molecules have specific
CC activity greater than that of human factor VIII and equal to or greater
CC than that of porcine factor VIII.
XX
SQ Sequence 7493 BP; 2487 A; 1503 C; 1436 G; 2067 T; 0 other;
Query Match 0.8%; Score 19; DB 19; Length 7493;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2347 AAAGAGTAATCCTTCA 2365
Db 206 AAAGAGTAATCCTTCA 224
RESULT 84
AAV91164
ID AAV91164 standard; cDNA to mRNA; 7493 BP.
XX
AC AAV91164;
XX
DT 26-NOV-1999 (first entry)
XX
DE Mouse factor VIII protein encoding nucleotide sequence.
XX
XX Factor VIII protein; hybrid protein; porcine; mouse; immunogenicity;
XX antigenic; procoagulant; factor X activation; haemophilia; human; ds.
XX
XX Mus musculus.
XX
XX W09946274-A1.
FN
XX 16-SEP-1999.
PD
XX 10-MAR-1999; 99WO-US05193.
PE
XX 10-MAR-1998; 98US-0037601.
PR
XX (UYEM-) UNIV EMORY.
XX
XX Loliar JS;
XX WPI; 1999-551355/46.
DR P-PSDB; AAV31596.
XX
XX New porcine and modified human factor VIII proteins for treating
PT hemophilia -
XX
XX Examples; Page 131-136; 187pp; English.
XX The invention provides DNA encoding porcine factor VIII, hybrid porcine

CC /human factor VIII, or modified human factor VIII having reduced
CC immunogenicity. Active factor VIII increases catalytic efficiency of
CC factor IXa towards factor X activation. The factor VIII proteins of the
CC invention can be produced by standard recombinant methodology and have
CC less antigenic activity, or greater procoagulant activity, than prior art
CC factor VIII. The invention is used to treat haemophilia. The present
CC sequence represents the nucleotide sequence encoding the A and C domains
CC of mouse factor VIII protein.
XX
SQ Sequence 7493 BP; 2487 A; 1503 C; 1436 G; 2067 T; 0 other;
Query Match 0.8%; Score 19; DB 20; Length 7493;
Best Local Similarity 100.0%; Pred. NO. 76;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2347 AAAGAGTAATCCTCA 2365
Db 206 AAAGAGTAATCCTCA 224
RESULT 85
ABK49586/C
ID ABK49586 standard; DNA; 143306 BP.
XX
XX ABK49586;
AC
XX
DT 15-JUL-2002 (first entry)
XX
DE Human transporter protein gene.
XX
XX Human; ds; gene; transporter; transgenic; transporter mediated disease;
KW drug screening; pharmacogenomic analysis; chromosome 18; SNP;
KW single nucleotide polymorphism.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH 3000..140308 /tag= a
FT CDS /product= "Transporter"
FT 3000..3051
FT exon /tag= b
FT /number= 1
FT intron 3052..62251
FT /tag= c
FT /number= 1
FT exon 62251..62601
FT /tag= d
FT /number= 2
FT intron 62602..84982
FT /tag= e
FT /number= 2
FT exon 84983..85100
FT /tag= f
FT /number= 3
FT intron 85101..98275
FT /tag= g
FT /number= 3
FT exon 98276..98436
FT /tag= h
FT /number= 4
FT intron 98437..133431
FT /tag= i
FT /number= 4
FT exon 133432..133618
FT /tag= j
FT /number= 5
FT intron 133619..137732
FT /tag= k
FT /number= 5
FT exon 137733..137875
FT /tag= l
FT /number= 6

FT intron 137876..139687
FT /tag= m
FT /number= 6
FT exon 139688..140305
FT /tag= n
FT /number= 7
FT variation replace (981,C)
FT /tag= o
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (1012,T)
FT /tag= p
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (1390,G)
FT /tag= q
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (3432,T)
FT /tag= r
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (4658,C)
FT /tag= s
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (4772,A)
FT /tag= t
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (7738,C)
FT /tag= u
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (9118,A)
FT /tag= v
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (9192,G)
FT /tag= w
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (9222,A)
FT /tag= x
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (9734,T)
FT /tag= y
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (9795,G)
FT /tag= z
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (9853,C)
FT /tag= aa
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (10397,A)
FT /tag= ab
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (11198,T)
FT /tag= ac
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (11470,T)
FT /tag= ad
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (11534,C)
FT /tag= ae
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (11557,G)
FT /tag= af
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (11540,G)
FT /tag= ag
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (11708,C)
FT /tag= ah
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (11953,T)
FT /tag= ai
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (11967,C)
FT /tag= aj
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (12172,C)

```

FT      /*tag= ak
FT      /standard name="Single-nucleotide polymorphism"
FT      replace (I2188,A)
FT      /*tag= al
FT      /standard name="Single-nucleotide polymorphism"
FT      replace (I3471..I3472,TTT)
FT      /*tag= am
FT      /standard name="Single-nucleotide polymorphism"
FT      replace (I3781,C)
FT      /*tag= an
FT      /standard name="Single-nucleotide polymorphism"
FT      replace (I4031,G)
FT      /*tag= ao
FT      /standard name="Single-nucleotide polymorphism"
FT      replace (I4110,C)
FT      /*tag= ap
FT      /standard name="Single-nucleotide polymorphism"
FT      replace (I4248,G or T)
FT      /*tag= aq
FT      /standard name="Single-nucleotide polymorphism"
FT      replace (I4676,C or A)
FT      /*tag= ar
FT      /standard name="Single-nucleotide polymorphism"
FT      replace (I4981,T)
FT      /*tag= as
FT      /standard name="Single-nucleotide polymorphism"
FT      replace (I6365,G)
FT      /*tag= at
FT      /standard name="Single-nucleotide polymorphism"
FT      replace (I6827,A)
FT      /*tag= au
FT      /standard name="Single-nucleotide polymorphism"
FT      replace (I7599,T)
FT      /*tag= av
FT      /standard name="Single-nucleotide polymorphism"
FT      replace (I8093,G)
FT      /*tag= aw
FT      /standard name="Single-nucleotide polymorphism"
FT      replace (I8226,C)
FT      /*tag= ax
FT      /standard name="Single-nucleotide polymorphism"
FT      replace (I8467..I8468,GTT or GGT)
FT      /*tag= ay
FT      /standard name="Single-nucleotide polymorphism"
FT      replace (I8511..I8512,TAC or TTC)
FT      /*tag= az
FT      /standard name="Single-nucleotide polymorphism"
FT      replace (I9107,A)
FT      /*tag= ba
FT      /standard name="Single-nucleotide polymorphism"
FT      replace (I9223,G)
FT      /*tag= bb
FT      /standard name="Single-nucleotide polymorphism"
FT      replace (I20102,C)
FT      /*tag= bc
FT      /standard name="Single-nucleotide polymorphism"
FT      replace (I21946,T)
FT      /*tag= bd
FT      /standard name="Single-nucleotide polymorphism"
FT      replace (I2240,C)
FT      /*tag= be
FT      /standard name="Single-nucleotide polymorphism"
FT      replace (I22679,G)
FT      /*tag= bf
FT      /standard name="Single-nucleotide polymorphism"
FT      replace (I23178,C or G)
FT      /*tag= bg
FT      /standard name="Single-nucleotide polymorphism"
FT      replace (I23370,G)
FT      /*tag= bh
FT      /standard name="Single-nucleotide polymorphism"
FT      replace (I23583,G)
FT      /*tag= bi

```

```

FT      /standard name="Single-nucleotide polymorphism"
FT      replace (I24159,A)
FT      /*tag= bj
FT      Query Match
FT      Best Local Similarity 100.0%; Pred. No. 85;
FT      Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy      321 TCAAGTAGATGCAAAATA 339
Db      18852 TCAAGTAGATGCAAAATA 18834
RESULT 86
AAA77296/C
ID AAA77296 standard; cDNA; 51 BP.
XX
AC AAA77296;
XX
DT 16-NOV-2000 (first entry)
XX
DE Human clone cg43997710 polymorphic site, SEQ ID NO:979.
XX
KW Human; single nucleotide polymorphism; SNP;
KW detection; identification; gene therapy; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT variation replace (26,A)
FT /*tag= a
XX
PM MO200029623-A2.
XX
PD 25-MAY-2000.
XX
PF 17-NOV-1999; 99WO-US27293.
XX
PR 17-NOV-1998; 98US-0109024.
PR 16-NOV-1999; 99US-0109024.
XX
PA (CURA-) CURAGEN CORP.
PI Shinketsu RA, Leach MD;
XX
DR WPI: 2000-387826/33.
XX
PT Human nucleic acids containing single nucleotide polymorphisms, useful
PT for treating a subject suffering, or at risk from a pathology due to
PT the presence of a sequence polymorphism -
XX
PS Claim 1; Page 453; 543pp; English.
XX
CC Sequences AAA76318-A77509 represent 1192 human nucleic acid sequences
CC which contain single nucleotide polymorphisms (SNPs). Sequences 1 to
CC 1112 (AAA76318-A77429) are consecutive pairs of nucleotides which
CC contain silent SNPs. Sequences 1113 to 1192 (AAA77430-A77509) are
CC consecutive pairs of nucleotides containing SNPs which result in changes
CC in the corresponding amino acid sequences (AAB11749-B11828). The SNPs in
CC sequences 1113 to 1128 (AAA77430-A77445) lead to conservative amino acid
CC changes, while those in sequences 1129 to 1186 (AAA77446-A77503) result
CC in non-conservative changes. The SNPs in sequences 1187 to 1192
CC (AAA77504-A77509) generate frameshift mutations. The invention also
CC relates to a method of detecting a polymorphic site in a nucleic acid and
CC a method of determining the relatedness of two nucleic acids. It also
CC encompasses peptides containing polymorphic sites, antibodies raised
CC against such peptides, and a method of detecting polymorphic
CC proteins/peptides using the antibodies. The nucleic acids are useful for
CC gene therapy of an individual having, suspected of having, or at risk of
CC developing a pathological condition due to the presence of a sequence
CC polymorphism. Such treatment would comprise administration of the
CC wild-type nucleic acid sequence. Antibodies raised against polymorphic
CC peptides can also be used in the treatment of such individuals.

```

XX Sequence 51 BP; 19 A; 6 C; 13 G; 13 T; 0 other;
 SQ Query Match 0.8%; Score 18; DB 21; Length 51;
 Best Local Similarity 100.0%; Pred. No. 2e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 1914 ATATATTAATTTGCTT 1931
 41 ATATATTAATTTGCTT 24
 Db
 RESULT 87
 ABX55959/c
 ID ABX55959 standard; cDNA; 324 BP.
 XX
 XX ABX55959;
 AC
 XX 26-FEB-2003 (first entry)
 DT
 XX Bovine EST associated with lactation/muscle/fat deposition #5888.
 DE
 XX Bovine; ss; EST; expressed sequence tag; lactation; LMPD;
 KM muscle deposition; fat deposition; genome mapping; gene identification;
 KW gene analysis; cattle breeding.
 XX
 XX Bos Taurus.
 OS
 XX US200217160-A1.
 PN
 XX 26-SEP-2002.
 PD
 XX 26-OCT-2001; 2001US-0983965.
 XX
 PF 17-DEC-1998; 98US-113678P.
 PR 15-DEC-1999; 99US-0465231.
 XX
 PA (BYAT/) BYAT J C.
 PA (MATH/) MATHIALAGAN N.
 PA (TAON/) TAO N.
 PA (WARR/) WARREN W C.
 XX
 PI Byatt JC; Mathialagan N; Tao N; Warren WC;
 DR WPI; 2003-102386/09.
 XX
 PT Purified nucleic acid molecules, useful for genome mapping, gene
 PT identification and analysis, cattle breeding or preparation of
 PT constructs for cattle gene expression and genetically improved cattle -
 XX
 XX Claim 2; SEQ ID No 5888; 38pp; English.
 PS
 XX The invention relates to a purified nucleic acid molecule associated with
 CC lactation or muscle and fat deposition (designated LMPD), derived
 CC from cattle, and the LMPD nucleic acid can specifically hybridize to a
 CC second nucleic acid molecule comprising any of 5912 nucleotide
 CC sequences, appearing as ABX50072-ABX55983, or complements of them.
 CC Also included are; (1) a transformed cell having a nucleic acid
 CC comprising an LMPD nucleic acid linked to a promoter and a 3' non-
 CC translated sequence and that functions in the cell to cause termination of
 CC transcription and addition of polyadenylated ribonucleotides to a 3' end
 CC of the mRNA molecule; and (2) determining a level or pattern of a
 CC molecule in a bovine cell or tissue comprising: (a) incubating a marker
 CC nucleic acid (comprising any of the 5912 nucleic acid sequences or its
 CC complement or fragment) with a complementary nucleic acid molecule
 CC obtained from the bovine cell or tissue, where hybridization between the
 CC marker nucleic acid and the complementary nucleic acid permits the
 CC detection of the molecule; and (b) detecting the level or pattern of the
 CC complementary nucleic acid, where the detection of the complementary
 CC nucleic acid is predictive of the level or pattern of the molecule.
 CC The LMPD nucleic acid is used for determining a level or pattern
 CC of a molecule in a bovine cell or tissue. It is useful for genome
 CC mapping, gene identification and analysis, cattle breeding, preparation

CC of constructs for use in cattle gene expression, or for genetically
 CC improving cattle. The present sequence is one of the 5912 bovine
 CC LMPD EST (expressed sequence tag) nucleic acids.
 CC Note: The present sequence was not shown in the specification but
 CC was obtained in electronic format from the USPTO web site:
 CC seqdata.uspto.gov/sequence.html?docid=20020137160.
 XX
 XX Sequence 324 BP; 113 A; 62 C; 49 G; 100 T; 0 other;
 SQ Query Match 0.8%; Score 18; DB 25; Length 324;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 1101 TTTAAATAGACTCAAAATT 1118
 108 TTTAAATAGACTCAAAATT 91
 Db
 RESULT 88
 AAC05686/c
 ID AAC05686 standard; cDNA; 327 BP.
 XX
 XX AAC05686;
 AC
 XX 06-OCT-2000 (first entry)
 DT
 XX Human secreted protein 5' EST, SEQ ID NO: 9761.
 DE
 XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KM gene therapy; chromosome mapping; ss.
 KW Homo sapiens.
 OS
 XX EP1033401-A2.
 PN
 XX 06-SEP-2000.
 PD
 XX 21-FEB-2000; 2000EP-0200610.
 PF
 XX 26-FEB-1999; 99US-0122487.
 PR
 XX (GSEST) GENSET.
 XX
 PA Dumas Milne Edwards J; Duclert A; Giordano J;
 PA
 XX WPI; 2000-500381/45.
 DR
 XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 XX Claim 1; SEQ ID 9761; 71pp + CD-ROM; English.
 PS
 XX The present sequence is one of a large number of 5' ESTs derived from
 CC cDNA encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.
 XX
 XX Sequence 327 BP; 79 A; 72 C; 70 G; 106 T; 0 other;
 SQ Query Match 0.8%; Score 18; DB 21; Length 327;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 228 TATCATCAGTGAGAAATT 245

RESULT 91
AA184590
ID AA184590 standard; cDNA; 379 BP.
XX
XX AA184590;
XX
XX
DT 06-NOV-2001 (first entry)
XX
XX Human polynucleotide SEQ ID NO 4650.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.
XX
XX Homo sapiens.
OS
XX WO200164835-A2.
XX
XX 07-SEP-2001.
XX
XX 26-FEB-2001; 2001WO-US04927.
XX
XX 28-FEB-2000; 2000US-0515126.
XX 18-MAY-2000; 2000US-0577409.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
PI
XX
XX WPI: 2001-514838/56.
XX P-PSDB; AA004659.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
XX Claim 1; SEQ ID NO 4650; 1359pp + Sequence Listing; English.
XX
XX The invention relates to human polynucleotides (AA179941-AA193841) and
CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activity/inhibit activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 379 BP; 160 A; 65 C; 68 G; 86 T; 0 other;

Query Match 0.8%; Score 18; DB 22; Length 379;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 751 CACAGCAACAAACAGCAA 768
DB 74 CACAGCAACAAACAGCAA 91

RESULT 92
AAC00113
ID AAC00113 standard; cDNA; 423 BP.
XX
XX AAC00113;

XX
XX 06-OCT-2000 (first entry)
DT
XX
XX Human secreted protein 5' EST, SEQ ID NO: 111.
DE
XX
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
XX Homo sapiens.
OS
XX EP1033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 99US-0122487.
XX
XX (GSEST) GENSEST.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
PI
XX
XX WPI: 2000-500381/45.
XX P-PSDB; AAC00107.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 1; SEQ ID 111; 71pp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC cDNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
XX Sequence 423 BP; 108 A; 91 C; 133 G; 89 T; 2 other;

Query Match 0.8%; Score 18; DB 21; Length 423;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2195 AAAAGAGCAAGATTG 2212
DB 130 AAAAGAGCAAGATTG 147

RESULT 93
AAC77174
ID AAC77174 standard; cDNA; 448 BP.
XX
XX AAC77174;
AC
XX
XX 08-FEB-2001 (first entry)
DT
XX
XX Human ORF2729 polynucleotide sequence SEQ ID NO:5457.
DE
XX
XX Human; open reading frame; ORF; detection; cytosolic; hepatotropic;
KW vulnery; antiporatic; antiparkinsonian; nocotropic; neuroprotective;
KW anticonvulsant; osteoplastic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antineumatic; antithyroid;

KW anti-nausea; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; anti-inflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000WO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shimkets RA, Leach M;
 XX
 DR WPI: 2000-602362/57.
 DR P-PSDB; AAB42965.
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 5; Page 4639; 5507pp; English.
 XX
 CC AACT4446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORF open reading frames 1 to 3161. The ORF
 CC sequences have activities such as: cytostatic; hepatocytic; vulnery;
 CC antiproliferative; antiparkinsonian; nootropic; neuroprotective;
 CC osteoprotic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatologic; immunosuppressive;
 CC anti-inflammatory; antibacterial; antiviral; antifungal; antineumatic;
 CC antithyroid; and anti-nausea. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORF-associated disorder. The
 CC nucleic acids can be used to express ORF proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorder, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, anti-inflammatory disease, to enhance
 CC coagulation, to inhibit thrombosis, and as a contraceptive.
 XX
 SQ Sequence 448 BP; 111 A; 106 C; 129 G; 102 T; 0 other;
 Query Match 0.8%; Score 18; DB 21; Length 448;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1914 ATATATTTAAATTTGCTT 1931
 DB 308 ATATATTTAAATTTGCTT 325
 RESULT 94
 AABN5206/C
 ID AABN5206 standard; cDNA; 495 BP.
 XX
 AC AABN5206;

XX
 DT 28-JUN-2002 (first entry)
 XX
 DE Human cancer related polynucleotide SEQ ID NO 5173.
 XX
 KW Human; cytostatic; gene expression; gene mapping; tissue profiling;
 KW gene therapy; cancer; tumour; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO20024500-A2.
 XX
 PD 21-FEB-2002.
 XX
 PF 16-AUG-2001; 2001WO-US25840.
 XX
 PR 16-AUG-2000; 2000US-226326P.
 XX
 PA (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.
 XX
 PI Escobedo J, Garcia PD, Sudduth-Klinger J, Reinhard C, Randazzo F;
 PI Lamson G, Scott EM, Zhang G, Kassam A, Pot D, Labat I;
 XX
 DR WPI: 2002-241905/29.
 XX
 PT New nucleic acid for producing a polypeptide, detecting differentially
 PT expressed genes correlated with a cancerous state of a mammalian cell,
 PT and inhibiting tumor growth -
 XX
 PS Claim 1; SEQ ID NO 5173; 883pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated polynucleotide (ABN27253-ABN33262)
 CC with cytostatic activity. The polynucleotide is used to produce a
 CC polypeptide, to detect differentially expressed genes correlated with a
 CC cancerous state of a mammalian cell and to inhibit tumour growth. The
 CC polynucleotide is used as a probe in mapping and tissue profiling. The
 CC encoded polypeptide and antibodies to the polypeptide can also be used
 CC for therapeutic and diagnostic purposes. The polynucleotide is useful for
 CC gene therapy.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences.
 XX
 SQ Sequence 495 BP; 157 A; 83 C; 87 G; 168 T; 0 other;
 Query Match 0.8%; Score 18; DB 24; Length 495;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 789 AAGTAATGACATGATA 806
 DB 312 AAGTAATGACATGATA 295
 RESULT 95
 AAC57418/C
 ID AAC57418 standard; DNA; 499 BP.
 XX
 AC AAC57418;
 XX
 DT 25-JAN-2001 (first entry)
 XX
 DE Archidonic acid metabolism related genomic biallelic marker #52.
 XX
 KW Human; biallelic marker; archidonic acid metabolism; genotyping;
 KW detection; hybridisation; phenotype; haplotype; SNP; polymorphic base;
 KW single nucleotide polymorphism; hybridisation assay; sequencing assay;
 KW specific amplification assay; identification; ERBM; 12-LO-RBM;
 KW eicosanoid-related biallelic marker; 12-LO-related biallelic marker; ds.
 XX
 OS Homo sapiens.
 XX

PN WO200047771-A2.
XX
XX 17-AUG-2000.
XX
XX 11-FEB-2000; 2000WO-IB00184.
XX
XX 12-FEB-1999; 99US-0119917.
PR 23-MAR-1999; 99US-0275267.
PR 07-MAY-1999; 99US-0133200.
XX
PA (GEST) GENSET.
XX Blumenfeld M, Bougueleret L, Chumakov I;
XX WPI; 2000-571881/53.
XX
XX Novel biallelic markers useful for detecting conditions and genotypes
PT associated with arachidonic acid metabolism -
XX
XX Claim 13; Page 274; 802pp; English.
XX
XX The present invention describes polynucleotides including biallelic
CC markers derived from genes involved in arachidonic acid metabolism and
CC from genomic regions flanking those genes. Methods from the present
CC invention may be used to select individuals for clinical trials and
CC predict responses to treatment with drugs. The polynucleotides may be
CC used in hybridisation assays, sequencing assays and specific
CC amplification assays for identifying an eicosanoid-related biallelic
CC marker (ERBM) or 12-LO-related biallelic marker, and for amplifying a
CC segment of nucleotide containing an ERBM. The polynucleotides are
CC useful in diagnostic kits. The markers may be used to detect conditions
CC and genotypes associated with arachidonic acid metabolism. AAC57367 to
CC AAC58018 and AAB24019 and AAB24020 represent sequences used in the
CC exemplification of the present invention.
CC N.B. Polymorphic bases (single nucleotide polymorphisms also known as
CC SNPs) in the polynucleotide sequences from the present invention have
CC been given as their corresponding degenerate bases e.g. a polymorphic
CC base of C or T has been given as Y.
XX
XX Sequence 499 BP; 123 A; 61 C; 86 G; 227 T; 2 other;
SQ
Query Match 0.8%; Score 18; DB 21; Length 499;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 751 CACAGCAACACACAGCAA 768
DB 76 CACAGCAACACACAGCAA 59
RESULT 96
AAC57938/c
ID AAC57938 standard; DNA; 499 BP.
XX
XX AAC57938;
AC
XX
XX 25-JAN-2001 (first entry)
DT
XX
XX Arachidonic acid metabolism related genomic biallelic marker #572.
DE
XX
XX Human; biallelic marker; arachidonic acid metabolism; genotyping;
KW detection; hybridisation; phenotype; SNP; polymorphic base;
KW single nucleotide polymorphism; hybridisation assay; sequencing assay;
KW specific amplification assay; identification; ERBM; 12-LO-RBM;
KW eicosanoid-related biallelic marker; 12-LO-related biallelic marker; ds.
XX
XX Homo sapiens.
OS
XX
XX WO200047771-A2.
PN
XX 17-AUG-2000.
PD
XX
XX 11-FEB-2000; 2000WO-IB00184.
PF

XX 12-FEB-1999; 99US-0119917.
PR 23-MAR-1999; 99US-0275267.
PR 07-MAY-1999; 99US-0133200.
XX
XX (GEST) GENSET.
XX Blumenfeld M, Bougueleret L, Chumakov I;
XX WPI; 2000-571881/53.
XX
XX Novel biallelic markers useful for detecting conditions and genotypes
PT associated with arachidonic acid metabolism -
XX
XX Claim 13; Page 709; 802pp; English.
XX
XX The present invention describes polynucleotides including biallelic
CC markers derived from genes involved in arachidonic acid metabolism and
CC from genomic regions flanking those genes. Methods from the present
CC invention may be used to select individuals for clinical trials and
CC predict responses to treatment with drugs. The polynucleotides may be
CC used in hybridisation assays, sequencing assays and specific
CC amplification assays for identifying an eicosanoid-related biallelic
CC marker (ERBM) or 12-LO-related biallelic marker, and for amplifying a
CC segment of nucleotides containing an ERBM. The polynucleotides are
CC useful in diagnostic kits. The markers may be used to detect conditions
CC and genotypes associated with arachidonic acid metabolism. AAC57367 to
CC AAC58018 and AAB24019 and AAB24020 represent sequences used in the
CC exemplification of the present invention.
CC N.B. Polymorphic bases (single nucleotide polymorphisms also known as
CC SNPs) in the polynucleotide sequences from the present invention have
CC been given as their corresponding degenerate bases e.g. a polymorphic
CC base of C or T has been given as Y.
XX
XX Sequence 499 BP; 123 A; 61 C; 86 G; 228 T; 1 other;
SQ
Query Match 0.8%; Score 18; DB 21; Length 499;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 751 CACAGCAACACACAGCAA 768
DB 76 CACAGCAACACACAGCAA 59
RESULT 97
AAL45031
ID AAL45031 standard; cDNA; 914 BP.
XX
XX AAL45031;
AC
XX
XX 16-MAY-2002 (first entry)
DT
XX
XX A thaliana AMP deaminase coding sequence fragment EST.
DE
XX
XX AMP deaminase; adenosine monophosphate; transgenic plant; EST;
KW herbicide resistance; herbicide; inhibitor; expressed sequence tag; ss.
XX
XX Arabidopsis thaliana.
OS
XX
XX Key Location/Qualifiers
FH CDS 2..880
FT /*tag= a
FT /product= "AMP deaminase fragment"
PN
XX
XX WO200206319-A2.
PD
XX 24-JAN-2002.
PD
XX 06-JUL-2001; 2001WO-EP07767.
PF
XX 17-JUL-2000; 2000DE-1035084.
PR
XX

PA (AVET) AVENTIS CROPS SCIENCE GMBH.
 XX
 XX Schulz A, Streiber W, Hanke C, Schmidt F, Schubel A;
 PI
 XX MPI, 2002-195802/25.
 DR
 XX P-PSDB; AAOI6944.
 XX
 XX New nucleic acid for plant adenosine monophosphate deaminase, useful in
 PT screening for herbicides and preparing herbicide-resistant plants -
 XX
 XX Example 2; Page 23; 51pp; German.
 PS
 XX The present invention provides the protein and coding sequences of the
 CC Arabidopsis thaliana adenosine monophosphate (AMP) deaminase. The coding
 CC sequence can be used to transform prokaryotic or eukaryotic cells,
 CC especially to produce transgenic plants (e.g. barley, rice, soya etc.)
 CC resistant to herbicidal inhibitors of AMP-deaminase, for recombinant
 CC production of proteins with AMP-deaminase activity, and to identify
 CC related genes in other organisms. The protein can be used for
 CC identification and biochemical/structural characterisation of new
 CC AMP-deaminase inhibitors and potential herbicides. The present sequence
 CC is an EST encoding a fragment of the protein of the invention.
 CC
 SQ Sequence 914 BP; 250 A; 215 C; 196 G; 253 T; 0 other;
 XX
 Query Match 0.8%; Score 18; DB 24; Length 914;
 Best Local Similarity 100.0%; Pred. No. 2,3e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1661 CACTGATTGGAAGAT 1678
 DB 716 CACTGATTGGAAGAT 733
 RESULT 98
 AAS56173/c
 ID AAS56173 standard; DNA; 924 BP.
 XX
 XX AAS56173;
 AC
 XX 13-FEB-2002 (first entry)
 DT
 XX Salmonella typhi DNA for cellular proliferation protein #206.
 DE
 XX Antisense; ds; prokaryotic cellular proliferation gene;
 KW antibiotic; antibacterial; drug design.
 XX
 XX Salmonella typhi.
 OS
 XX WO200170955-A2.
 PN
 XX 27-SEP-2001.
 PD
 XX 21-MAR-2001; 2001WO-US09180.
 PF
 XX 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207272P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX
 XX (ELIT-) ELITRA PHARM INC.
 PA
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GU;
 PI Yamamoto RT, Xu HH;
 XX
 XX MPI: 2001-611495/70.
 DR P-PSDB; AAU38314.
 XX
 XX New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -

XX
 XX Claim 27; Seq ID No 9810; 51pp; English.
 PS
 XX The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence encodes an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pcr_sequences.
 CC
 SQ Sequence 924 BP; 241 A; 213 C; 269 G; 200 T; 1 other;
 XX
 Query Match 0.8%; Score 18; DB 23; Length 924;
 Best Local Similarity 100.0%; Pred. No. 2,3e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 2106 ACTTCAAGCGGATGAAG 2123
 DB 854 ACTTCAAGCGGATGAAG 837
 RESULT 99
 ABL12667/c
 ID ABL12667 standard; CDNA; 1038 BP.
 XX
 XX ABL12667;
 AC
 XX 26-MAR-2002 (first entry)
 DT
 XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 32483.
 DE
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX
 XX Drosophila melanogaster.
 OS
 XX WO200171042-A2.
 PN
 XX 27-SEP-2001.
 PD
 XX 23-MAR-2001; 2001WO-US09231.
 PF
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 PR
 XX (PEKE) PE CORP NY.
 PA
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI MPI; 2001-656860/75.
 DR P-PSDB; ABB68564.
 XX
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 interactions -
 XX
 XX Claim 1; SEQ ID NO 32483; 21pp + Sequence Listing; English.
 PS
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB55737-AB572072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1038 BP; 295 A; 238 C; 222 G; 283 T; 0 other;

Query Match 0.8%; Score 18; DB 23; Length 1038;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1882 TAATTGATTATTCCTCA 1899
Db 788 TAATTGATTATTCCTCA 771

RESULT 100
AAC44226/C
ID AAC44226 standard; DNA; 1077 BP.
XX
AC AAC44226;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 42087.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
FN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131445.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144337.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.

PR 02-AUG-1999; 99US-0146388.
 PR C2-AUG-1999; 99US-0146389.
 PR C3-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.

PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 0.8%; Score 18; DB 21; Length 1077;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1498 GCGACTTGGCAACCAAA 1515
 Db 518 GCGACTTGGCAACCAAA 501

Search completed: November 14, 2003, 03:59:40
 Job time : 683 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 03:15:31 ; Search time 4993 Seconds

(without alignments)
11628.959 Million cell updates/sec

Title: US-09-765-271-55

Perfect score: 2389
Sequence: 1 TTCTTACGAGTTGGACTGT.....TAAGTAGCAAAAATAAAC 2389

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size : 0

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 200 summaries

Database :

EST:*
1: em_estba:*
2: em_estba:*
3: em_estba:*
4: em_estba:*
5: em_estba:*
6: em_estba:*
7: em_estba:*
8: em_estba:*
9: em_estba:*
10: em_estba:*
11: em_estba:*
12: em_estba:*
13: em_estba:*
14: em_estba:*
15: em_estba:*
16: em_estba:*
17: em_estba:*
18: em_estba:*
19: em_estba:*
20: em_estba:*
21: em_estba:*
22: em_estba:*
23: em_estba:*
24: em_estba:*
25: em_estba:*
26: em_estba:*
27: em_estba:*
28: em_estba:*
29: em_estba:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Match	Length	DB	ID	Description
1	22	0.9	261	10	BE528097	BE528097 M70N065TM	
2	22	0.9	618	9	AU236995	AU236995 AU236995	
3	22	0.9	682	28	BH970231	BH970231 cdf81d11.	
4	21	0.9	411	13	BU473013	BU473013 603365019	

5	21	0.9	499	12	B1544663	B1544663 603242619
6	21	0.9	549	10	AW932212	AW932212 EST358055
7	21	0.9	581	13	B3397983	B3397983 603534550
8	21	0.9	640	28	AZ391393	AZ391393 1M0153A03
9	21	0.9	664	13	B3345704	B3345704 604170839
10	21	0.9	684	29	AG084591	AG084591 Pan t10g1
11	21	0.9	690	12	B3549886	B3549886 BU549886
12	21	0.9	731	13	B3403931	B3403931 604141302
13	21	0.9	742	13	B2834994	B2834994 60362178
14	21	0.9	754	13	B1154421	B1154421 603140883
15	21	0.9	769	13	B211482	B211482 604159635
16	21	0.9	799	13	B3088927	B3088927 603538542
17	21	0.9	870	29	CC088871	CC088871 CSU-K33r
18	21	0.9	935	13	B273390	B273390 603531527
19	20	0.8	175	10	B538572	B538572 BS38572
20	20	0.8	228	28	AZ502341	AZ502341 1M0341023
21	20	0.8	233	28	AZ769909	AZ769909 1M0571001
22	20	0.8	267	28	A2943045	A2943045 2M0203306
23	20	0.8	371	10	B8803034	B8803034 BB803034
24	20	0.8	406	28	AQ046037	AQ046037 RPI11-36
25	20	0.8	471	28	AQ216823	AQ216823 HS 3253 B
26	20	0.8	496	29	BZ685217	BZ685217 PUBCP45TD
27	20	0.8	540	9	AW239782	AW239782 P11c.PX0
28	20	0.8	584	28	BH369198	BH369198 AG-ND-139
29	20	0.8	588	28	BH766941	BH766941 BMBC345E
30	20	0.8	595	14	CA370410	CA370410 650577 NC
31	20	0.8	603	13	BW282910	BW282910 BM282910
32	20	0.8	622	28	AZ524614	AZ524614 234PBD06
33	20	0.8	630	28	AQ412913	AQ412913 RPI1-11-1
34	20	0.8	636	28	BH765160	BH765160 BMBC354A
35	20	0.8	643	13	BW112810	BW112810 BM112810
36	20	0.8	649	13	BW243771	BW243771 BM243771
37	20	0.8	651	10	BG698682	BG698682 602703065
38	20	0.8	656	10	BP296373	BP296373 036PBD02
39	20	0.8	656	28	BH085526	BH085526 RPI1-24-2
40	20	0.8	658	28	AZ373971	AZ373971 1M0126P17
41	20	0.8	664	13	BW261677	BW261677 BM261677
42	20	0.8	668	12	BG911162	BG911162 602806563
43	20	0.8	689	9	AJ453442	AJ453442 AJ453442
44	20	0.8	689	13	BW262397	BW262397 BW262397
45	20	0.8	698	10	BF294340	BF294340 004PBA08
46	20	0.8	706	10	BS527573	BS527573 BS527573
47	20	0.8	708	13	BW027074	BW027074 BM027074
48	20	0.8	712	9	AV862261	AV862261 AV862261
49	20	0.8	713	14	CA763329	CA763329 BR060002B
50	20	0.8	736	29	CNS01FBA	AL141617 Anopheles
51	20	0.8	748	13	BW132597	BW132597 BM132597
52	20	0.8	761	29	BZ995097	BZ995097 PUDFB72TD
53	20	0.8	765	13	BW131850	BW131850 BM131850
54	20	0.8	770	29	CNS01MFC	AL150861 Anopheles
55	20	0.8	776	14	CB598673	CB598673 AGENCOURT
56	20	0.8	792	10	BE130716	BE130716 L48-922T3
57	20	0.8	795	28	BH374713	BH374713 AG-ND-161
58	20	0.8	809	12	B3566263	B3566263 BU566263
59	20	0.8	812	29	CC159651	CC159651 1903C07.9
60	20	0.8	831	28	BH045505	BH045505 RPI1-24-3
61	20	0.8	876	29	CNS07DNK	AL1440594 T3 end of
62	20	0.8	923	13	BG03498	BG03498 602301896
63	20	0.8	957	10	BQ712408	BQ712408 AGENCOURT
64	20	0.8	2185	11	AK037932	AK037932 Mus muscu
65	19	0.8	66	28	BH907885	BH907885 SALK 0446
66	19	0.8	110	29	CC021347	CC021347 359T_1.24
67	19	0.8	136	14	H74951	H74951 539 Random
68	19	0.8	175	10	BE505229	BE505229 dc19G09.x
69	19	0.8	186	28	BH867080	BH867080 hg93b06.y
70	19	0.8	220	9	AV236691	AV236691 AV236691
71	19	0.8	257	28	AO948125	AO948125 Sheared D
72	19	0.8	271	10	BA402033	BA402033 BSMC011XC
73	19	0.8	296	14	CD265572	CD265572 QVO-OT003
74	19	0.8	303	10	AM880404	AM880404 K0824E07-
75	19	0.8	303	12	BM290438	BM290438 BM290438
76	19	0.8	304	10	BM290438	BM290438 BM290438
77	19	0.8	304	10	BM367558	BM367558 BM367558

```

78      19      0.8      309      29      B2244064      B2244064      CH230-362
79      19      0.8      317      10      B3138766      B3138766
80      19      0.8      319      10      BB463922      BB463922
81      19      0.8      324      9      AV046714      AV046714
82      19      0.8      333      9      A1547552      A1547552
83      19      0.8      335      28      BH294683      BH294683
84      19      0.8      345      9      AA558478      AA558478
85      19      0.8      360      28      A0825612      A0825612
86      19      0.8      364      28      A0910031      A0910031
87      19      0.8      370      10      BF374552      BF374552
88      19      0.8      375      10      BF376477      BF376477
89      19      0.8      377      9      A1883328      A1883328
90      19      0.8      386      13      BY649148      BY649148
91      19      0.8      396      10      BF376465      BF376465
92      19      0.8      401      10      BE002498      BE002498
93      19      0.8      401      29      B2941299      B2941299
94      19      0.8      402      10      BF885054      BF885054
95      19      0.8      409      10      BF554284      BF554284
96      19      0.8      412      13      BY605266      BY605266
97      19      0.8      414      9      AL961018      AL961018
98      19      0.8      415      12      BM248963      BM248963
99      19      0.8      417      10      BE505361      BE505361
100     19      0.8      418      13      BY457007      BY457007
101     19      0.8      434      28      A0207842      A0207842
102     19      0.8      434      29      CC029019      CC029019
103     19      0.8      440      10      BE635488      BE635488
104     19      0.8      440      28      A0188923      A0188923
105     19      0.8      448      9      AM046397      AM046397
106     19      0.8      457      28      BS6992      BS6992
107     19      0.8      459      9      A0081605      A0081605
108     19      0.8      462      28      A0393398      A0393398
109     19      0.8      467      9      A1502258      A1502258
110     19      0.8      479      13      BY492908      BY492908
111     19      0.8      485      10      BF942976      BF942976
112     19      0.8      491      28      A0944295      A0944295
113     19      0.8      495      9      AL867799      AL867799
114     19      0.8      499      13      BU663360      BU663360
115     19      0.8      501      28      A2481176      A2481176
116     19      0.8      514      10      BF025045      BF025045
117     19      0.8      517      28      BH788839      BH788839
118     19      0.8      524      9      AL651149      AL651149
119     19      0.8      525      9      AM743009      AM743009
120     19      0.8      536      9      AL644915      AL644915
121     19      0.8      536      28      A2364035      A2364035
122     19      0.8      537      10      BE098306      BE098306
123     19      0.8      539      14      CB822845      CB822845
124     19      0.8      541      9      A1009609      A1009609
125     19      0.8      546      29      CC172145      CC172145
126     19      0.8      547      9      AL598888      AL598888
127     19      0.8      553      28      A2396203      A2396203
128     19      0.8      557      9      AL639919      AL639919
129     19      0.8      560      29      B2821447      B2821447
130     19      0.8      561      28      A2494994      A2494994
131     19      0.8      566      28      A2231630      A2231630
132     19      0.8      573      28      A0489684      A0489684
133     19      0.8      576      14      CB867995      CB867995
134     19      0.8      577      28      BH786618      BH786618
135     19      0.8      578      10      BF007111      BF007111
136     19      0.8      582      28      BH398547      BH398547
137     19      0.8      583      28      A2338047      A2338047
138     19      0.8      589      28      BS9270      BS9270
139     19      0.8      595      13      BQ286841      BQ286841
140     19      0.8      597      9      AUI00630      AUI00630
141     19      0.8      597      12      BM491868      BM491868
142     19      0.8      609      13      B0479735      B0479735
143     19      0.8      615      28      BS5701      BS5701
144     19      0.8      620      9      A1574440      A1574440
145     19      0.8      620      9      AL647248      AL647248
146     19      0.8      629      9      AL792122      AL792122
147     19      0.8      630      9      AL874461      AL874461
148     19      0.8      631      9      AL627924      AL627924
149     19      0.8      634      9      AL891387      AL891387
150     19      0.8      636      9      AL642283      AL642283

```

```

151     19      0.8      636      28      A2411549      A2411549
152     19      0.8      640      9      AL873932      AL873932
153     19      0.8      642      10      BB467078      BB467078
154     19      0.8      643      9      AL954908      AL954908
155     19      0.8      643      9      AM391405      AM391405
156     19      0.8      648      13      BQ859663      BQ859663
157     19      0.8      649      9      AL861345      AL861345
158     19      0.8      651      9      AL893908      AL893908
159     19      0.8      652      9      AL888054      AL888054
160     19      0.8      652      13      BQ397518      BQ397518
161     19      0.8      652      14      CA814289      CA814289
162     19      0.8      656      9      AL973266      AL973266
163     19      0.8      657      29      CC445648      CC445648
164     19      0.8      658      14      CD213529      CD213529
165     19      0.8      661      13      BU269133      BU269133
166     19      0.8      662      9      AL857690      AL857690
167     19      0.8      664      28      A2154788      A2154788
168     19      0.8      668      9      AL886201      AL886201
169     19      0.8      672      29      B2530407      B2530407
170     19      0.8      673      14      CB670071      CB670071
171     19      0.8      675      28      BH942657      BH942657
172     19      0.8      675      29      AG041704      AG041704
173     19      0.8      678      29      AG093187      AG093187
174     19      0.8      682      29      AG103329      AG103329
175     19      0.8      689      13      BU466818      BU466818
176     19      0.8      689      28      A2356132      A2356132
177     19      0.8      690      13      BU467071      BU467071
178     19      0.8      696      12      BM436991      BM436991
179     19      0.8      696      29      CNS0105C      CNS0105C
180     19      0.8      701      9      AL962927      AL962927
181     19      0.8      701      29      B2786641      B2786641
182     19      0.8      707      28      A2104118      A2104118
183     19      0.8      711      13      BU004246      BU004246
184     19      0.8      711      14      CA750429      CA750429
185     19      0.8      714      13      BU450047      BU450047
186     19      0.8      716      10      BB340597      BB340597
187     19      0.8      716      13      BU285148      BU285148
188     19      0.8      719      9      AL632000      AL632000
189     19      0.8      721      13      BQ868729      BQ868729
190     19      0.8      726      14      CB657006      CB657006
191     19      0.8      726      28      BH931736      BH931736
192     19      0.8      727      28      B2057551      B2057551
193     19      0.8      728      13      BU481748      BU481748
194     19      0.8      732      28      BH028155      BH028155
195     19      0.8      734      13      BU725729      BU725729
196     19      0.8      737      14      CB527334      CB527334
197     19      0.8      739      14      CB650155      CB650155
198     19      0.8      749      29      AG075433      AG075433
199     19      0.8      750      14      CB655195      CB655195
200     19      0.8      753      13      BU329419      BU329419

```

ALIGNMENTS

```

RESULT 1      261 bp      mRNA      linear      EST 19-MAR-2001
LOCUS      BE528097
DEFINITION  WTON06STM Arabidopsis developing seed Arabidopsis thaliana cDNA
ACCESSION   BE528097
VERSION     BE528097.1
KEYWORDS    GI:9786075
SOURCE      Arabidopsis thaliana (thale cress)
ORGANISM   Arabidopsis thaliana

```

```

REFERENCE   1 (bases 1 to 261)
AUTHORS    White,J.A., Todd,J., Newman,T., Focks,N., Gire,T., Martinez de
            Iriarte,O., Jaworski,J.G., Ohlrogge,J. and Bennett,C.
TITLE      A new set of Arabidopsis expressed sequence tags from developing
            seeds. The metabolic pathway from carbohydrates to seed oil

```


JOURNAL	Plant Physiol. 124 (4), 1582-1594 (2000)
MEDLINE	20567808
PUBMED	1115876
COMMENT	Contact: Benning, C

```

RESULT 4
LOCUS      BU473013/c      411 bp      mRNA      linear      EST 30-NOV-2002
DEFINITION 603365019p1 CSEQRBN21 Gallus gallus cDNA clone CHEST262f5 5', mRNA
sequence.
ACCESSION  BU473013
VERSION     BU473013.1  GI:25966590
KEYWORDS
SOURCE      Gallus gallus (chicken)
ORGANISM    Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianine; Gallus.
REFERENCE   1 (bases 1 to 411)
AUTHORS    Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
            Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
TITLE      A Comprehensive Collection of Chicken cDNAs
JOURNAL    Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE    22355534
PUBMED     12445392
COMMENT     Contact: Simon Hubbard
            Department of Biomolecular Sciences
            University of Manchester Institute of Science and Technology (UMIST)
            PO Box 88, Manchester, M60 1QD, UK
            Tel: 01612008930
            Fax: 01612360409
            Email: Simon.Hubbard@umist.ac.uk.
            Location/Qualifiers
FEATURES
source
1..411
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Layer"
/db_xref="taxon:9031"
/clone="CHEST262f5"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_id="CSEQRBN21"
/note="Organ: ovary; Vector: pluescript II KS(+); Site 1:
EcoRI; Site 2: NotI; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was bluntended, ligated to NotI adapters, digested with EcoRI
, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
(1996) : 791, except that a significantly longer
reannealing hybridization was used."
BASE COUNT  105 a      65 c      94 g      146 t      1 others
ORIGIN
Query Match      0.9%; Score 21; DB 13; Length 411;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy      2190 AACCTAAGAGCAGAGCTT 2210
|||||
Db      273 AACCTAAGAGCAGAGCTT 253
|||||
RESULT 5
LOCUS      B1544663      499 bp      mRNA      linear      EST 05-SEP-2001
DEFINITION 603242619p1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5284997 5',
mRNA sequence.
ACCESSION  B1544663
ORIGIN
Query Match      0.9%; Score 21; DB 13; Length 411;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy      2190 AACCTAAGAGCAGAGCTT 2210
|||||
Db      273 AACCTAAGAGCAGAGCTT 253
|||||

```

```

VERSION 1
KEYWORDS    B1544663.1  GI:15431975
SOURCE      EST.
ORGANISM    Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE   1 (bases 1 to 499)
AUTHORS    NIH-MGC http://mgs.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgabs-remail.nih.gov
            Tissue Procurement: Miklos Palokvits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNLN at:
            http://image.llnl.gov
            Plate: L1AM11720 row: f column: 06
            High quality sequence stop: 299.
            Location/Qualifiers
FEATURES
source
1..499
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5284997"
/tissue_type="hippocampus"
/lab_host="DH10B"
/clone_id="NIH_MGC_95"
/note="Organ: brain; Vector: pBluescript (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (GCGAG
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.5 kb and
normalized to R0T 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT  122 a      142 c      165 g      70 t
ORIGIN
Query Match      0.9%; Score 21; DB 12; Length 499;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy      389 GAGGAATCAATCGACAAA 409
|||||
Db      414 GAGGAATCAATCGACAAA 434
|||||
RESULT 6
LOCUS      AM932212      549 bp      mRNA      linear      EST 18-MAY-2001
DEFINITION EST358055 tomato fruit mature green, TMU Lycopersicon esculentum
cDNA clone CLEF471J3 5', mRNA sequence.
ACCESSION  AM932212
VERSION     AM932212.1  GI:8107613
KEYWORDS
SOURCE      Lycopersicon esculentum (tomato)
ORGANISM    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1 (bases 1 to 549)
AUTHORS    Alcalá,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F.,
            Upson,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Roming,C.M.,
            Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
            Generation of ESTs from tomato fruit tissue
            Unpublished
            Contact: CUGI
            Clemson University Genomics Institute

```

FEATURES
source
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.
Location/Qualifiers

1. .549
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/cultivar="TA96"
/db_xref="taxon:4081"
/clone="cLEF47L13"
/tissue_type="fruit pericarp"
/dev_stage="mature green (3-5 days pre-ripening)"
/lab_host="SOLR"
/clone_lib="tomato fruit mature green, TAMU"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; cLEF Fruit were tagged at the 1cm stage and harvested 3-5 days prior to ripening. Fruit were cut in half to verify the seeds were indeed 'immature' and the seeds and locules were discarded prior to freezing the pericarp"
BASE COUNT 166 a 97 c 124 g 161 t 1 others
ORIGIN

Query Match 0.9%; Score 21; DB 10; Length 549;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 306 AGGCTGATATGTTATCAAG 326
DB 135 AGGCTGATATGTTATCAAG 155

RESULT 7
BU397983/c 581 bp mRNA linear EST 27-NOV-2002
LOCUS 603534550F1 CSBQCHN58 Gallus gallus cDNA CHEST494c13 5', mRNA
DEFINITION sequence.
ACCESSION BU397983
VERSION BU397983.1 GI:25767039
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
1 (bases 1 to 581)
Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E., Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
A Comprehensive Collection of Chicken CDNA's
Curr. Biol. 12 (22), 1965-1969 (2002)
COMMENT
PUBLISHED 22335534
12445392
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk
Location/Qualifiers

FEATURES
source
1. .581
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Compton Line 151"
/db_xref="taxon:9031"
/clone="CHEST494c13"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="CSBQCHN58"
/note="Organ: small intestine; Vector: pBluescript II KS(+

); Site_1: EcoRI; Site_2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was bluntended, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."
BASE COUNT 156 a 98 c 119 g 208 t
ORIGIN

Query Match 0.9%; Score 21; DB 13; Length 581;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2190 AACTCAAGACGACGAGTTT 2210
DB 368 AACTCAAGACGACGAGTTT 348

RESULT 8
AZ391393/c 640 bp DNA linear GSS 03-OCT-2000
LOCUS 1M0153A03R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC1M0153A03 R, genomic survey sequence.
ACCESSION AZ391393
VERSION AZ391393.1 GI:10506436
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 640)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamli, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A. and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
TITLE Unpublished
JOURNAL
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 1000 Std Error: 0.00
Plate: 0153 row: A column: 03
Seq primer: CACACAGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 640.
Location/Qualifiers

FEATURES
source
1. .640
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0153A03"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
<http://www.jax.org/resources/documents/dnares/>. The DNA was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 215 a 139 c 183 g 102 t 1 others
ORIGIN

Query Match 0.9%; Score 21; DB 28; Length 640;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1539 ACGAAGTCGATTCGCTCAAT 1559
DB 332 ACGAAGTCGATTCGCTCAAT 312

RESULT 9
LOCUS BU345704 664 bp mRNA linear EST 28-NOV-2002
DEFINITION 60417083991 CSEQCHN68 Gallus gallus cDNA clone CHEST1033m10 5',
ACCESSION BU345704
VERSION BU345704.1 GI:25853705
KEYWORDS EST
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 664)
AUTHORS Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
TITLE A Comprehensive Collection of Chicken CDNAs
JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE 22335534
PUBMED 12445392
COMMENT Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

FEATURES

source 1..664
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Compton Line 151"
/db_xref="taxon:9031"
/clone="CHEST1033m10"
/sex="Female"
/tissue_type="cerebellum"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="CSEQCHN68"
/note="Organ: brain; Vector: pluscript II KS(+); Site 1:
EcoRI; Site 2: NotI; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunted, ligated to NotI adaptors, digested with EcoRI

, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pluscript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996) : 791, except that a significantly longer reannealing hybridization was used."

BASE COUNT 186 a 97 c 144 g 237 t
ORIGIN

Query Match 0.9%; Score 21; DB 13; Length 664;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2190 AACTCAAGAGCAGAGTTT 2210
DB 295 AACTCAAGAGCAGAGTTT 275

RESULT 10
LOCUS AG084591 681 bp DNA linear GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-082604.F, genomic survey sequence.
ACCESSION AG084591
VERSION AG084591.1 GI:1636393
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.

REFERENCE 1
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Torok, Y., Watanabe, H. and Sakaki, Y.
TITLE BAC end sequences of library PTB
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 681)
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Torok, Y., Watanabe, H. and Sakaki, Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Aao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimbases@bic.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the Rad process and may have higher chance of
clone tracking errors.

PRIMERS

Sequencing: -21M13
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
Location/Qualifiers

FEATURES

source 1..681
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-082604.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 125 a 213 c 191 g 152 t
ORIGIN

Query Match 0.9%; Score 21; DB 29; Length 681;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1755 CAGATGTTAAGCAATCCAA 1775
DB 142 CAGATGTTAAGCAATCCAA 122

RESULT 11
 BJ549886
 LOCUS
 DEFINITION BJ549886 K. Sato unpublished cDNA library, cv. Haruna Nijo
 germination shoots Hordeum vulgare subsp. vulgare cDNA clone
 bag935c04 3', mRNA sequence.
 ACCESSION BJ549886
 VERSION BJ549886
 KEYWORDS EST.
 SOURCE BJ549886.1 GI:24968336
 ORGANISM
 Hordeum vulgare subsp. vulgare
 Hordeum vulgare subsp. vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
 ; Triticeae; Hordeum.
 1 (bases 1 to 690)
 Sato, K., Saitoh, D. and Takeda, K.
 Barley EST sequencing project in NIG and Okayama Univ
 Unpublished
 Contact: Tadao Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshini@genes.nig.ac.jp.
 Location/Qualifiers
 1..690
 /organism="Hordeum vulgare subsp. vulgare"
 /mol_type="mRNA"
 /cultivar="Haruna Nijo"
 /db_xref="taxon:112509"
 /clone="bag935c04"
 /tissue_type="shoots"
 /dev_stage="germination"
 /clone_1ib="K. Sato unpublished cDNA library, cv. Haruna
 Nijo germination shoots"
 BASE COUNT 239 a 154 c 134 g 163 t
 ORIGIN

Query Match 0.9%; Score 21; DB 12; Length 690;
 Best Local Similarity 100.0%; Pred. No. 89;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1078 AGCCCGCACTGCACCAA 1098
 ||||||||||||||||||
 Db 76 AGCCCGCACTGCACCAA 96

RESULT 12
 BU403931/c
 LOCUS BU403931 731 bp mRNA linear EST 27-NOV-2002
 DEFINITION 60414102P1 CSEQCHN59 Gallus gallus cDNA clone CHEST974d2 5', mRNA
 sequence.
 ACCESSION BU403931
 VERSION BU403931
 KEYWORDS EST.
 SOURCE BU403931.1 GI:25772987
 ORGANISM
 Gallus gallus (chicken)
 Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianidae; Gallus.
 1 (bases 1 to 731)
 Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
 A Comprehensive Collection of Chicken cDNAs
 Curr. Biol. 12 (22), 1965-1969 (2002)
 22335534
 12445392
 MEDLINE
 PUBMED
 COMMENT Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology (UMIST)

PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.
 Location/Qualifiers
 1..731
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="White Leghorn, Hisex"
 /db_xref="taxon:9831"
 /clone="CHEST974d2"
 /dev_stage="36"
 /lab_host="DH10B"
 /clone_1ib="CSEQCHN59"
 /note="Organ: limbs; Vector: pluescript II KS(+); Site_1:
 EcoRI, Site_2: NotI. This normalized library was
 constructed from 1 million independent clones. cDNA
 synthesis was initiated using an oligo(dT) primer, using
 methylated C in the first strand synthesis reaction.
 Following this first strand reaction, double-stranded cDNA
 was blunt-ended, ligated to NotI adapters, digested with EcoRI
 , size-selected, and cloned into the NotI and EcoRI
 compatible sites of a custom modified MCS of the
 pluescript (KS+) vector. The library was normalized in 2
 rounds using conditions adapted from Soares et al., PNAS
 (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
 (1996) 791, except that a significantly longer
 reannealing hybridization was used."
 BASE COUNT 200 a 121 c 157 g 252 t
 ORIGIN

Query Match 0.9%; Score 21; DB 13; Length 731;
 Best Local Similarity 100.0%; Pred. No. 90;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2190 AACTCAAGAGACAGAGTTT 2210
 ||||||||||||||||||
 Db 531 AACTCAAGAGACAGAGTTT 511

RESULT 13
 BU283494/c
 LOCUS BU283494 742 bp mRNA linear EST 27-NOV-2002
 DEFINITION 603862178P1 CSEQCHN54 Gallus gallus cDNA clone CHEST875k20 5', mRNA
 sequence.
 ACCESSION BU283494
 VERSION BU283494
 KEYWORDS EST.
 SOURCE BU283494.1 GI:25732950
 ORGANISM
 Gallus gallus (chicken)
 Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianidae; Gallus.
 1 (bases 1 to 742)
 Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
 A Comprehensive Collection of Chicken cDNAs
 Curr. Biol. 12 (22), 1965-1969 (2002)
 22335534
 12445392
 MEDLINE
 PUBMED
 COMMENT Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology (UMIST)

PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.
 Location/Qualifiers
 1..742
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="Compton Line 151"

```

/db_xref="taxon:9031"
/clone="CHEST875K20"
/sex="Female"
/tissue_type="not cerebrum or cerebellum"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="CSEQCHN54"
/note="Organ: brain; Vector: pBluescript II KS(+); Site_1:
EcoRI; Site_2: NotI; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunt-ended, ligated to NotI adapters, digested with EcoRI
, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
(1996): 791, except that a significantly longer
reannealing hybridization was used."
BASE COUNT      201 a      124 c      159 g      258 t
ORIGIN
Query Match      0.9%; Score 21; DB 13; Length 742;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      2190 AACTCAAGAGCAGAGATT 2210
|||||
Db      359 AACTCAAGAGCAGAGATT 339

RESULT 14
LOCUS      BU115421      754 bp      mRNA      linear      EST 25-NOV-2002
DEFINITION      60314083F1 CSEQCHL15 Gallus gallus cDNA clone CHEST132e22 5', mRNA
sequence.
ACCESSION      BU115421
VERSION      BU115421.1 GI:25321521
KEYWORDS      EST.
SOURCE      Gallus gallus (chicken)
ORGANISM      Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE      1 (bases 1 to 754)
AUTHORS      Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
TITLE      A Comprehensive Collection of Chicken cDNAs
JOURNAL      Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE      22335534
PUBMED      12445392
COMMENT      Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST
)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

FEATURES
source
1..754
location/Qualifiers
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Compton line 151"
/db_xref="taxon:9031"
/clone="CHEST132e22"
/sex="Female"
/tissue_type="cerebrum"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="CSEQCHL15"

```

```

/note="Organ: brain; Vector: pBluescript II KS(+); Site_1:
EcoRI; Site_2: NotI; Modification of pBluescript II KS(+)
[Stratagene] vector to accommodate cDNA produced with the
T-trimmed protocol (Construction of uni-directionally
cloned cDNA libraries from messenger RNA for improved 3'
end DNA sequencing by Glenn Fu, et al. U.S. Patent # 6,387
,624). Cut pBluescript II KS(+) with NotI and EcoRI.
Ligate in double stranded adaptor containing BglI and
BamHI sites [5'ggccgctgcagcccccggatccgaaanaag]
[5'aattcttttcgcatccgggtgcagc]"
BASE COUNT      192 a      135 c      150 g      277 t
ORIGIN
Query Match      0.9%; Score 21; DB 13; Length 754;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      2190 AACTCAAGAGCAGAGATT 2210
|||||
Db      667 AACTCAAGAGCAGAGATT 647

RESULT 15
LOCUS      BU211482      769 bp      mRNA      linear      EST 25-NOV-2002
DEFINITION      60415963F1 CSEQCHN03 Gallus gallus cDNA clone CHEST1013e10 5',
mRNA sequence.
ACCESSION      BU211482
VERSION      BU211482.1 GI:25385632
KEYWORDS      EST.
SOURCE      Gallus gallus (chicken)
ORGANISM      Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE      1 (bases 1 to 769)
AUTHORS      Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
TITLE      A Comprehensive Collection of Chicken cDNAs
JOURNAL      Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE      22335534
PUBMED      12445392
COMMENT      Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST
)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

FEATURES
source
1..769
location/Qualifiers
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hixex"
/db_xref="taxon:9031"
/clone="CHEST1013e10"
/tissue_type="whole embryo"
/dev_stage="20-21"
/lab_host="DH10B"
/clone_lib="CSEQCHN03"
/note="Organ: whole embryo; Vector: pBluescript II KS(+);
Site_1: EcoRI; Site_2: NotI; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunt-ended, ligated to NotI adapters, digested with EcoRI
, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6

```

BASE COUNT 257 a 177 c 147 g 188 t
 (1996): 791, except that a significantly longer
 reannealing hybridization was used."

Query Match 0.9%; Score 21; DB 13; Length 769;
 Best Local Similarity 100.0%; Pred. No. 90;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2190 AACTCAAGACGACGAGTTT 2210
 |||||
 DB 389 AACTCAAGACGACGAGTTT 409

RESULT 16
 BU08927/c 799 bp mRNA linear EST 26-NOV-2002
 LOCUS 603538542F1 CSEQCHN60 Gallus gallus cDNA clone CHEST504c11 5', mRNA
 DEFINITION sequence.

ACCESSION BU08927
 VERSION BU08927.1 GI:25816928
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus

REFERENCE Boardman, P. E., Sanz-Ezquerro, J., Overton, I. M., Burt, D. W., Bosch, E.,
 Fong, W. T., Tickle, C., Brown, W. R. A., Wilson, S. A. and Hubbard, S. J.
 A Comprehensive Collection of Chicken CDNAS
 Curr. Biol. 12 (22), 1965-1969 (2002)

AUTHORS Boardman, P. E., Sanz-Ezquerro, J., Overton, I. M., Burt, D. W., Bosch, E.,
 Fong, W. T., Tickle, C., Brown, W. R. A., Wilson, S. A. and Hubbard, S. J.
 TITLE A Comprehensive Collection of Chicken CDNAS
 JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
 MEDLINE 22335534
 PUBMED 12445392

COMMENT Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology (UMIST)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.
 Location/Qualifiers
 1..799

FEATURES
 source

/organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="Compton Line 151"
 /db_xref="taxon:9031"
 /clone="CHEST504c11"
 /sex="Female"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_1lb="CSEQCHN60"
 /note="Organ: heart; Vector: pBluescript II KS(+); Site_1:
 EcoRI; Site_2: NotI; This normalized library was
 constructed from 1 million independent clones. cDNA
 synthesis was initiated using an oligo(dT) primer, using
 methylated C in the first strand synthesis reaction.
 Following this first strand reaction, double-stranded cDNA
 was blunt-ended, ligated to NotI adapters, digested with EcoRI
 , size-selected, and cloned into the NotI and EcoRI
 compatible sites of a custom modified MCS of the
 pBluescript (KS+) vector. The library was normalized in 2
 rounds using conditions adapted from Soares et al., PNAS
 (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
 (1996): 791, except that a significantly longer
 reannealing hybridization was used."

BASE COUNT 221 a 134 c 176 g 268 t

Query Match 0.9%; Score 21; DB 13; Length 799;
 Best Local Similarity 100.0%; Pred. No. 91;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2190 AACTCAAGACGACGAGTTT 2210
 |||||
 DB 465 AACTCAAGACGACGAGTTT 445

RESULT 17
 CC088871 870 bp DNA linear GSS 16-APR-2003
 LOCUS CC088871
 DEFINITION CSU-K33r.8E23.SP6 CSU-K33r Aedes aegypti genomic clone
 CSU-K33r.8E23, genomic survey sequence.

ACCESSION CC088871
 VERSION CC088871.1 GI:29944500
 KEYWORDS GSS.
 SOURCE Aedes aegypti (yellow fever mosquito)
 ORGANISM Aedes aegypti

REFERENCE Loftus, B., Shetty, J., Severson, D., Brown, S. and Knudson, D.
 1 (bases 1 to 870)
 Title: End sequencing of Aedes aegypti BACS
 Unpublished
 Other_GSSs: CSU-K33r.8E23.T7
 Contact: Brendan Loftus
 Department of Eukaryotic Genomics
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-838-3543
 Fax: 301-838-0208
 Email: ente@tigr.org
 Library was provided by Susan Brown and Dennis Knudson at Colorado
 State University.
 Seq primer: SP6
 Class: BAC ends.

FEATURES
 source

/organism="Aedes aegypti"
 /mol_type="genomic DNA"
 /strain="Rexville"
 /db_xref="taxon:7159"
 /clone="CSU-K33r.8E23"
 /clone_1lb="CSU-K33r"
 /note="Vector: pBlotBAC11; Site_1: HindIII"

BASE COUNT 276 a 161 c 161 g 272 t

Query Match 0.9%; Score 21; DB 29; Length 870;
 Best Local Similarity 100.0%; Pred. No. 92;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1222 TGTAAATCTTGAAGCA 1242
 |||||
 DB 91 TGTAAATCTTGAAGCA 111

RESULT 18
 BU273390 935 bp mRNA linear EST 26-NOV-2002
 LOCUS BU273390
 DEFINITION 603531527F1 CSEQCHN53 Gallus gallus cDNA clone CHEST486f8 5', mRNA
 sequence.

ACCESSION BU273390
 VERSION BU273390.1 GI:25544340
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus

REFERENCE Boardman, P. E., Sanz-Ezquerro, J., Overton, I. M., Burt, D. W., Bosch, E.,
 Fong, W. T., Tickle, C., Brown, W. R. A., Wilson, S. A. and Hubbard, S. J.
 A Comprehensive Collection of Chicken CDNAS
 Curr. Biol. 12 (22), 1965-1969 (2002)

AUTHORS Boardman, P. E., Sanz-Ezquerro, J., Overton, I. M., Burt, D. W., Bosch, E.,
 Fong, W. T., Tickle, C., Brown, W. R. A., Wilson, S. A. and Hubbard, S. J.
 TITLE A Comprehensive Collection of Chicken CDNAS
 JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)

MEDLINE
22335534
12445392
PUBMED
COMMENT
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 10D, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
Location/Qualifiers

FEATURES
source
1. 935
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Compton Line 151"
/db_xref="taxon:9031"
/clone="CHEST486f8"
/sex="Female"
/tissue_type="cerebrum"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="CSECHN53"
/note="Organ: Brain; Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI; This normalized library was constructed from 1 million independent clones. CDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded CDNA was blunt-ended, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1996) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

BASE COUNT
247 a 184 c 205 g 299 t

ORIGIN

Query Match .0 9%; Score 21; DB 13; Length 935;
Best Local Similarity 100.0%; Pred.No. 94;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2190 AACTCAAGACGACGAGTTT 2210
|||||
500 AACTCAAGACGACGAGTTT 480

RESULT 19
BB538572 175 bp mRNA linear EST 31-JUL-2000
LOCUS BB538572 RIKEN full-length enriched, 0 day neonate eyeball Mus
DEFINITION musculus CDNA clone E130012112 3', mRNA sequence.
ACCESSION BB538572
VERSION BB538572.1 GI:9594072
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 175)
Komno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ichii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadoya, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, C., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugihara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomiwa, N., Toya, T., Tsunoda, Y., Watanabe, S., Yamamuta, T., Yamashita, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

TITLE
RIKEN Mouse ESTs (Komno, H., et al.)
JOURNAL
Unpublished
COMMENT
Contact: Yoshinide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermostabilization and thermoinactivation of the labile enzymes by trehalose and its application for the synthesis of full length CDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (12), 520-524 (1998)
Itoh, M., Katsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length CDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
Location/Qualifiers
1. 175
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="E130012112"
/tissue_type="eyeball"
/dev_stage="0 day neonate"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 0 day neonate eyeball"
/note="Site 1: SalI; Site 2: BamHI; CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand CDNA was primed with a primer [5'
GAGAGAGAGAGCGCGCCGACCTCGAGTTTCTTTTCTTTTNN 3']. CDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand CDNA was prepared with the primer adapter of sequence [5'
GAGAGAGAGATTCGAGTTTAATTAATTAATCCCCCCCCCC 3']. CDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLX I."

BASE COUNT
45 a 52 c 31 g 47 t

ORIGIN

Query Match 0.8%; Score 20; DB 10; Length 175;
Best Local Similarity 100.0%; Pred.No. 2,2e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 54 TTTCCTATATAGATGAAAA 73
|||||
Db 140 TTTCCTATATAGATGAAAA 159

RESULT 20
AZ502341/c 228 bp DNA linear GSS 05-OCT-2000
LOCUS AZ502341 Mouse 10kb plasmid UUC1M library Mus musculus genomic
DEFINITION clone UUC1M0341023 F, genomic survey sequence.
ACCESSION AZ502341
VERSION AZ502341.1 GI:10683573
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 228)

REFERENCE 1 (bases 1 to 228)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

TITLE Unpublished

JOURNAL Unpublished

COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0341 row: O column: 23
Seq primer: CGTTGTAAACGACGCGCAGT
Class: plasmid ends
High quality sequence stop: 228.

FEATURES
source
1..228
Location/Qualifiers
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUCG1M0341023"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUCG1M library"
/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 55 a 54 c 46 g 73 t

ORIGIN

Query Match 0.8%; Score 20; DB 28; Length 228;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 TTCTCATATAGATGGAAGA 73
|||||
Db 220 TTCTCATATAGATGGAAGA 201

RESULT 21 233 bp DNA linear GSS 16-FEB-2001
LOCUS AZ769909
DEFINITION IM0571J01F Mouse 10kb plasmid UUCG1M library Mus musculus genomic clone UUCG1M0571J01 F, genomic survey sequence.
ACCESSION AZ769909
VERSION AZ769909.1 GI:12890543
KEYWORDS GSS.

SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 233)

REFERENCE 1 (bases 1 to 233)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

TITLE Unpublished

JOURNAL Unpublished

COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0571 row: J column: 01
Seq primer: CGTTGTAAACGACGCGCAGT
Class: plasmid ends
High quality sequence stop: 233.

FEATURES
source
1..233
Location/Qualifiers
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUCG1M0571J01"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUCG1M library"
/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 57 a 48 c 44 g 84 t

ORIGIN

Query Match 0.8%; Score 20; DB 28; Length 233;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 318 TTATCAAGGTAGATGGAAGA 337
|||||
Db 73 TTATCAAGGTAGATGGAAGA 92

RESULT 22 267 bp DNA linear GSS 26-APR-2001
LOCUS AZ943045
DEFINITION ZM0203J06R Mouse 10kb plasmid UUCG2M library Mus musculus genomic clone UUCG2M0203J06 R, genomic survey sequence.
ACCESSION AZ943045
VERSION AZ943045.1 GI:13806817
KEYWORDS GSS.

KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 267)
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenan, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhuesern, A., and Wright, D., Weiss, R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0203 row: J column: 06
Seq primer: CACACGAGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 267.

FEATURES
SOURCE location/Qualifiers
1..267
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUCG2M020J06"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 87 a 68 c 56 g 56 t
ORIGIN

Query Match 0.8%; Score 20; DB 28; Length 267;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1711 AGCCTACTTAAGAAAAAG 1730
|||||
Db 71 AGCCTACTTAAGAAAAAG 90

RESULT 23
BB803034 371 bp mRNA linear EST 19-NOV-2001
LOCUS BB803034 RIKEN full-length enriched, 16 days neonate male
DEFINITION diencephalon Mus musculus cDNA clone G630051N24 3', mRNA sequence.
ACCESSION BB803034

VERSION BB803034.1 GI:16975663
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 371)
AUTHORS Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Imotani, K., Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Komno, H., Kouda, M., Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Nomura, R., Okazaki, T., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Watanishi, A., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.
TITLE RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al. 2001)
JOURNAL Unpublished
COMMENT Contact: Yoshinhide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-resgsc.riken.go.jp/
URL: <http://genome.gsc.riken.go.jp/>
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Iwata, M., Ohara, E., Watanishi, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, D., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A., and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Komno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y., and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.
e mouse tissues.

FEATURES
SOURCE location/Qualifiers
1..371
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="G630051N24"
/sex="male"
/tissue_type="dienecephalon"
/dev_stage="16 days neonate"
/clone_id="RIKEN full-length enriched, 16 days neonate male dienecephalon"

BASE COUNT 97 a 88 c 87 g 99 t
ORIGIN

Query Match 0.8%; Score 20; DB 10; Length 371;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 TTTCCTATATGATGAAAA 73
|||||
Db 200 TTTCCTATATGATGAAAA 219

RESULT 24
A0046037/c

LOCUS A0046037 406 bp DNA linear GSS 14-APR-1999
 DEFINITION RPCI11-36019.TK RPCI-11 Homo sapiens genomic clone RPCI-11-36019,
 genomic survey sequence.
 ACCESSION A0046037
 VERSION A0046037.1 GI:3314964
 KEYWORDS GSS.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 406)
 Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
 Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and
 Venter,J.C.
 Use of BAC End Sequences for Sequence-Ready Map Building (1998)
 TITLE Unpublished
 JOURNAL
 COMMENT Other_GSSs: RPCI11-36019.TU
 Contact: Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: mdadams@tigr.org
 Clones are derived from the human BAC library RPCI-11. For BAC
 library availability, please contact Pieter de Jong
 (pieterdejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
 Research Genetics (info@resgen.com). BAC end search page:
 http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
 Class: BAC ends.
 Location/Qualifiers
 1..406
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="GDB:7513794"
 /db_xref="taxon:9606"
 /clone="RPCI-11-36019"
 /sex="Male"
 /cell_type="lymphocytes"
 /clone_11b="RPCI-11"
 /note="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;
 RPCI11 Human Male BAC Library"
 BASE COUNT 113 a 64 c 63 g 166 t
 ORIGIN
 Query Match 0.8%; Score 20; DB 28; Length 406;
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1214 TCTGAACGTCTAAATCT 1233
 ||||||||||||||||||
 Db 199 TCTGAACGTCTAAATCT 180
 RESULT 25
 LOCUS A0216823 471 bp DNA linear GSS 19-SEP-1998
 DEFINITION HS_3253_B1_C01.T7 CIT Approved Human Genomic Sperm Library D Homo
 sapiens genomic clone Plate=3253 Col=1 Row=F, genomic survey
 sequence.
 ACCESSION A0216823
 VERSION A0216823.1 GI:3632143
 KEYWORDS GSS.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 471)
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.
 Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome
 Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 PUBMED 10449764
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Sequence Tagged Connector
 Plate: 3253 row: F column: 1
 Class: BAC ends
 High quality sequence stop: 471.
 Location/Qualifiers
 1..471
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="Plate=3253 Col=1 Row=F"
 /sex="male"
 /clone_11b="CIT Approved Human Genomic Sperm Library D"
 /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
 E-Coli DH10B"
 BASE COUNT 143 a 103 c 85 g 139 t 1 others
 ORIGIN
 Query Match 0.8%; Score 20; DB 28; Length 471;
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 310 TGGATATGTTATCAAGTAG 329
 ||||||||||||||||||
 Db 448 TGGATATGTTATCAAGTAG 429
 RESULT 26
 LOCUS BZ685217 496 bp DNA linear GSS 05-FEB-2003
 DEFINITION PUBCP45TD ZM.0.6.1.0 KB Zea mays genomic clone ZMBMTa024H18,
 genomic survey sequence.
 ACCESSION BZ685217
 VERSION BZ685217.1 GI:28244313
 KEYWORDS GSS.
 SOURCE Zea mays
 ORGANISM Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoidae; Andropogoneae; Zea.
 1 (bases 1 to 496)
 Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick
 A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
 Maize Genomics Consortium
 Unpublished
 Contact: Cathy Whitelaw
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-838-5843
 Fax: 301-838-0208
 Email: whitelaw@tigr.org
 Seq primer: TF
 Class: sheared ends.
 Location/Qualifiers
 1..496
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /strain="B73"
 /db_xref="taxon:4577"
 /clone="ZMBMTa024H18"
 /clone_11b="ZM.0.6.1.0 KB"
 /note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
 COT selected genomic DNA library"

ORIGIN	137 a	81 c	82 g	196 t
Query Match	0.8%;	Score 20;	DB 29;	Length 496;
Best Local Similarity	100.0%;	Pred. No. 2.7e+02;		
Matches	20;	Conservative	0;	Mismatches 0;
Indels	0;	Gaps	0;	
OY	1308	ACCAAGATTTCATGATAA	1327	
Db	33	ACCAAGATTTCATGATAA	14	
RESULT 27				
LOCUS	AM239782	540 bp	mRNA	linear
DEFINITION	ptl1c.pK002.h12 chicken MDV infected T cell cDNA library Gallus gallus cDNA clone ptl1c.pK002.h12 5' similar to (AB023214) KIAA0997 protein [Homo sapiens], mRNA sequence.			
ACCESSION	AM239782			
VERSION	AM239782.1	GI:6579522		
KEYWORDS	EST;			
SOURCE	Gallus gallus (chicken)			
ORGANISM	Gallus gallus			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauromia; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.			
REFERENCE	1 (bases 1 to 540)			
AUTHORS	Morgan,R.			
TITLE	Chicken ESTs from MDV infected T cells			
JOURNAL	Unpublished			
COMMENT	Contact: Robin Morgan			
	University of Delaware			
	Townsend Hall, Newark, DE 19717, USA			
	Tel: 302-831-1341			
	Fax: 302-831-2822			
	Email: morgan@udel.edu			
	Clones can be ordered online at http://www.chickest.udel.edu .			
FEATURES	Location/Qualifiers			
source	1..540			
	/organism="Gallus gallus"			
	/mol_type="mRNA"			
	/db_xref="taxon:9031"			
	/clone="ptl1c.pK002.h12"			
	/sex="Male and Female"			
	/cell_type="MDV infected splenic T cell"			
	/lab_host="E.coli DH5 alpha"			
	/clone_1ib="chicken MDV infected T cell cDNA library"			
	/note="Vector: pB42AD"			
BASE COUNT	172 a	120 c	101 g	146 t
ORIGIN				1 others
Query Match	0.8%;	Score 20;	DB 9;	Length 540;
Best Local Similarity	100.0%;	Pred. No. 2.7e+02;		
Matches	20;	Conservative	0;	Mismatches 0;
Indels	0;	Gaps	0;	
OY	751	CACAAGCAACAACGACACA	770	
Db	85	CACAAGCAACAACGACACA	104	
RESULT 28				
LOCUS	BH369198	584 bp	DNA	linear
DEFINITION	AG-ND-139K6.TF ND-TAM Anopheles gambiae genomic clone AG-ND-139K6, genomic survey sequence.			
ACCESSION	BH369198			
VERSION	BH369198.1	GI:17315299		
KEYWORDS	GSS;			
SOURCE	Anopheles gambiae (African malaria mosquito)			
ORGANISM	Anopheles gambiae			
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Preygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;			
	Anopheles.			

REFERENCE	1 (bases 1 to 584)
AUTHORS	Hong,Y.S., Hogan,J.R., Wang,X., Sarkar,A., Sim,C., Loftus,B.J., Ren, 'C', Huff,E.R., Carlile,J.L., Black,K., Zhang,H.-B., Gardner,M.J., and Collins,F.H.
TITLE	Construction of a BAC library and generation of BAC end sequence-tagged connectors for genome sequencing of the African malaria mosquito Anopheles gambiae
JOURNAL	Mol. Genet. Genomics 268 (6), 720-728 (2003)
MEDLINE	22542063
PUBMED	12655398
COMMENT	Other GSSs: AG-ND-139K6.TR Contact: Brendan J Loftus Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel.: 301 838 0208 Fax: 301 838 3543 Email: b.loftus@tigr.org This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest. Seq primer: M13 For Class: BAC ends. Location/Qualifiers 1..584 /organism="Anopheles gambiae" /mol_type="genomic DNA" /strain="PEST" /db_xref="taxon:7165" /clone="AG-ND-139K6" /clone_lib="ND-TAM" /note="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT	152 a 159 c 147 g 126 t
ORIGIN	
Query Match	0.8%, Score 20; DB 28; Length 584;
Best Local Similarity	100.0%; Pred. No. 2,7e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	2141 CTTGCTGAGCCGAGATCCC 2160
Db	581 CTTGCTGAGCCGAGATCCC 562
RESULT 29	
BH766941/c	588 bp DNA linear GSS 20-MAR-2002
LOCUS	
DEFINITION	BMDAC34580977 PSU Brugia malayi Genomic Bac Library 3 Brugia malayi
ACCESSION	BH766941
VERSION	BH766941.1
KEYWORDS	GI:19564705
SOURCE	GSS.
ORGANISM	Brugia malayi Brugia malayi Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
REFERENCE	Onchocercidae; Brugia.
AUTHORS	1 (bases 1 to 588) Whitton,C., Daud,J., Ware,J., Quail,M., Hall,N., Barrell,B., Fostere, 'J', Guiliano,D., Slack,B. and Blaxter,M.
TITLE	Genome survey sequences from the human parasitic nematode Brugia malayi
JOURNAL	Unpublished
COMMENT	Contact: Blaxter ML Institute of Cell, Animal and Population Biology University of Edinburgh Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9 3JT, UK

Tel: +44 131 650 6760
 Fax: +44 131 670 5450
 Email: mark.blaxter@ed.ac.uk
 Sequenced from the Brugia malayi BAC library constructed by Claire Whitton and Dr Mike Quail. The sequence was generated by The Pathogen Sequencing Unit, The Sanger Institute, Cambridge, UK in collaboration with Mark Blaxter, ICAPB, University of Edinburgh, Edinburgh, UK.
 Seg primer: T7 (TAATACGACTCACTATAGG)
 Class: BAC ends.

FEATURES

source

Location/Qualifiers
 1..588
 /organism="Brugia malayi"
 /mol_type="genomic DNA"
 /strain="T8S"
 /db_xref="taxon:6279"
 /sex="Mixed (male and female)"
 /tissue_type="whole parasite"
 /dev_stage="microfilaria (L1)"
 /clone_lib="Brugia malayi Genomic BAC Library 3"
 /note="Vector: pBAC3.6; Site 1: BamH I; Brugia malayi genomic DNA was partially cleaved with Sau3A I and size fractionated. 7,392 clones were generated with mean insert size ~48 kbp. The library was constructed by Claire Whitton, Blaxter Nematode Genetics Lab, University of Edinburgh, UK, and Dr Mike Quail, The Pathogen Sequencing Unit, The Sanger Centre, Cambridge, UK."

BASE COUNT 190 a 88 c 95 g 214 t 1 others

ORIGIN

Query Match 0.8%; Score 20; DB 28; Length 588;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1276 AACTGCTAAAAAGAAAATG 1295
 |||
 347 AACTGCTAAAAAGAAAATG 328

DB

347 AACTGCTAAAAAGAAAATG 328

RESULT 30

LOCUS

CA370410 595 bp mRNA linear EST 06-NOV-2002

DEFINITION

650577 NCCGWA 1RT Oncorhynchus mykiss cDNA clone 1RT22U21_B_E11 5', mRNA sequence.

ACCESSION

CA370410

VERSION

CA370410.1 GI:24683000

KEYWORDS

EST.
 Oncorhynchus mykiss (rainbow trout)

SOURCE

Oncorhynchus mykiss
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.

REFERENCE

1 (bases 1 to 595)
 Rexroad, C.E. and Keeler, J.W.

AUTHORS

Sequence analysis of a rainbow trout normalized cDNA library

TITLE

Unpublished

JOURNAL

Contact: Rexroad CE
 USDA, ARS, National Center for Cool and Cold Water Aquaculture
 11876 Leetown Road, Kearneysville, WV 25430, USA

COMMENT

Tel: 304 724 8340 x2129
 Fax: 304 725 0351

FEATURES

Email: crexroad@nccgwa.ars.usda.gov
 Single pass sequencing. Bases called with phred v0.020425.c and

SOURCE

Trimmed with the aid of the trim_alt option. Vector identified by
 cross_match v0.990329.
 Seg primer: AGCGGATACCAATTTCACACAGCA.

FEATURES

Location/Qualifiers
 1..595

source

/organism="Oncorhynchus mykiss"
 /mol_type="mRNA"
 /db_xref="taxon:8022"
 /clone="1RT22U21_B_E11"
 /tissue_type="pooled"

/lab_host="DH10B"
 /clone_lib="NCCGWA 1RT"
 /note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
 library made from pooled tissue from brain, gill, liver,
 spleen, muscle, and kidney."
 BASE COUNT 178 a 198 c 135 g 84 t

ORIGIN

Query Match 0.8%; Score 20; DB 14; Length 595;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

757 CACCAACGCAACACTAACA 776
 |||
 96 CACCAACGCAACACTAACA 115

DB

96 CACCAACGCAACACTAACA 115

RESULT 31

LOCUS

BM282910 603 bp mRNA linear EST 11-NOV-2002

DEFINITION

BM282910 Nori Satoh unpublished cDNA library, gonad cDNA
 intestinalis cDNA clone cigd020g09 5', mRNA sequence.

ACCESSION

BM282910

VERSION

BM282910.1 GI:24863521

KEYWORDS

EST.
 Ciona intestinalis

SOURCE

Ciona intestinalis
 Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
 Phlebobranchia; Clonidae; Ciona.

REFERENCE

1 (bases 1 to 603)
 Satou, Y., Shin, I. T., Kohara, Y. and Satoh, N.

AUTHORS

Expressed genes in Ciona intestinalis (2002c)

TITLE

Unpublished

JOURNAL

Contact: Nori Satoh
 Department of Zoology
 Kyoto University
 Sakyo-Ku, Kyoto, Kyoto 606-8502, Japan

COMMENT

Tel: 81-75-753-4081
 Fax: 81-75-705-1113
 Email: satoh@ascidian.zool.kyoto-u.ac.jp.

FEATURES

Location/Qualifiers
 1..603

source

/organism="Ciona intestinalis"
 /mol_type="mRNA"
 /db_xref="taxon:7719"
 /clone="cigd020g09"
 /tissue_type="gonad"
 /clone_lib="Nori Satoh unpublished cDNA library, gonad"

BASE COUNT

204 a 92 c 130 g 177 t

ORIGIN

Query Match 0.8%; Score 20; DB 13; Length 603;
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

2312 ATCATGCGAGAGCAAGAAA 2331
 |||
 234 ATCATGCGAGAGCAAGAAA 253

DB

234 ATCATGCGAGAGCAAGAAA 253

RESULT 32

LOCUS

AZ524614 622 bp DNA linear GSS 07-MAY-2001

DEFINITION

234pDb06 Pb MEN #21 Plasmodium berghei genomic 3', genomic survey
 sequence.

ACCESSION

AZ524614

VERSION

AZ524614.1 GI:13964669

KEYWORDS

GSS.
 Plasmodium berghei

SOURCE

Plasmodium berghei
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE

1 (bases 1 to 622)
 Carlton, J.M.-R. and Dame, J.B.

TITLE The Plasmodium vivax and P. berghei gene sequence tag projects
JOURNAL Parasitol. Today (Regul. Ed.) 16 (10), 409 (2000)
COMMENT Contact: Dame JB
Dept. of Pathobiology, College of Veterinary Medicine
University of Florida
2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
Tel: 352 392 4700
Fax: 352 392 9704
Email: damejb@mail.vetmed.ufl.edu
Seq primer: M13(-20) forward
Class: shotgun.

FEATURES
SOURCE Location/Qualifiers
1..622

/organism="Plasmodium berghei"
/mol_type="genomic DNA"
/strain="ANKA clone 15cyl (clone of the ANKA 8417 clone)"
/db_xref="taxon:5821"
/dev_stage="asexual blood forms"
/lab_host="Mus musculus"
/clone_1fb="Pb MBN #21"
/note="Vector: pBluescript SK(+) vector DNA, phagemid excised from lambda ZAP; Site_1: EcoRV; Site_2: EcoRV; genomic DNA was prepared from asynchronous blood stage forms of the cloned ANKA isolate of P. berghei grown in laboratory Swiss white mice. The DNA was purified from ultracentrifugation and precipitated. Purified DNA was digested with mung bean nuclease in the presence of 36-38% formamide at 50 C, as described (Verlick, K.D., Imberki, R.B., and McCutchan, T.F. 1988. Nucleic Acids Research 16:6883-6896). The ends of the digestion fragments were polished using T4 DNA polymerase, and the fragments size selected in the range 500-2000 bp. These were ligated into the EcoRV-cleaved and dephosphorylated pBluescript SK(+) vector. Recombinant plasmids were used to transform E. coli XL10-Gold host cells."

BASE COUNT 269 a 45 c 71 g 237 t
ORIGIN

Query Match 0.8%; Score 20; DB 28; Length 622;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1594 TTTGATGACATGATATTA 1613
DB 125 TTTGATGACATGATATTA 144

RESULT 33
LOCUS AQ412913 630 bp DNA linear GSS 23-MAR-1999
DEFINITION RPT-11-194P5.TV RPT-11 Homo sapiens genomic clone RPT-11-194P5,
genomic survey sequence.
ACCESSION AQ412913
VERSION AQ412913.1 GI:4472401
KEYWORDS GSS.

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 630)
Zhao, S., Adams, M.D., Niernan, W., Malik, J., de Jong, P. and Venter

AUTHORS J.C.

TITLE Use of BAC End Sequences from Library RPT-11 for Sequence-Ready
Map Building
JOURNAL Unpublished
COMMENT Contact: Shaying Zhao, William Niernan, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org

Clones are derived from the human BAC library RPT-11. For BAC library availability, please contact Pieter de Jong (pieter@tigr.org, med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genet cs (http://www.tigr.org/tcdb/humgen/bac_end_search/bac_end_search.html).
Seq primer: SP6
Class: BAC ends.

FEATURES
SOURCE Location/Qualifiers
1..630

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="GDB:7574476"
/db_xref="taxon:9606"
/clone="RPT-11-194P5"
/sex="Male"
/cell_type="Lymphocytes"
/clone_1fb="RPT-11"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; RPT111 Human Male BAC library"

BASE COUNT 123 a 94 c 108 g 305 t
ORIGIN

Query Match 0.8%; Score 20; DB 28; Length 630;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2087 AGTAGAGTCCAAATAGAA 2106
DB 254 AGTAGAGTCCAAATAGAA 235

RESULT 34
LOCUS BH765160 636 bp DNA linear GSS 20-MAR-2002
DEFINITION BMBAC354A03SP6.PSU Brugia malayi Genomic BAC Library 3 Brugia
malayi genomic, genomic survey sequence.
ACCESSION BH765160
VERSION BH765160.1 GI:19562924
KEYWORDS GSS.

SOURCE Brugia malayi
ORGANISM Brugia malayi
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Brugia.

REFERENCE 1 (bases 1 to 636)
Whitton, C., Daub, J., Ware, J., Quail, M., Hall, N., Barrell, B., Foster

AUTHORS J., Guillano, D., Slatko, B. and Blaxter, M.

TITLE Genome survey sequences from the human parasitic nematode Brugia
malayi

JOURNAL Unpublished
COMMENT Contact: Blaxter ML
Institute of Cell, Animal and Population Biology
University of Edinburgh
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
3JT, UK
Tel: +44 131 650 6760
Fax: +44 131 670 5450
Email: mark.blaxter@ed.ac.uk

Sequenced from the Brugia malayi BAC library constructed by Claire
Whitton and Dr Mike Quail. The sequence was generated by The
Pathogen Sequencing Unit, The Sanger Institute, Cambridge, UK in
collaboration with Mark Blaxter, ICAFB, University of Edinburgh,
Edinburgh, UK.

Seq primer: SP6 (ATTAGGTGACCTATAG)
Class: BAC ends.

FEATURES
SOURCE Location/Qualifiers
1..636

/organism="Brugia malayi"
/mol_type="genomic DNA"
/strain="IRS"
/db_xref="taxon:6279"
/sex="Mixed (male and female)"
/tissue_type="whole parasite"

/dev_stage="microfilaria (L1)"
/clone_1ib="Brugia malayi Genomic Bac Library 3"
/note="Vector: pBACE3.6; Site 1: BamH I; Brugia malayi
genomic DNA was partially cleaved with Sau3A I and size
fractionated. 7,392 clones were generated with mean insert
size ~48 kbp. The library was constructed by Claire
Whitton, Blaxter Nematode Genetics Lab, University of
Edinburgh, UK, and Dr Mike Quail, The Pathogen Sequencing
Unit, The Sanger Centre, Cambridge, UK."

BASE COUNT 226 a 90 c 96 g 224 t
ORIGIN

Query Match 0.8%; Score 20; DB 28; Length 636;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1741 TCCATCTCCAGACGACGATG 1760
|||||
Db 520 TCCATCTCCAGACGACGATG 501
|||||

RESULT 35
BM12810/c 643 bp mRNA linear EST 24-OCT-2002
LOCUS BM12810 Nori Satoh unpublished cDNA library, tailbud embryo Clona
DEFINITION intestinalis cDNA clone rc1b062j14 3', mRNA sequence.
ACCESSION BM12810
VERSION BM12810.1 GI:24359467
KEYWORDS EST.
SOURCE Clona intestinalis
ORGANISM Clona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
Phlebobranchia; Clonidae; Clona.
REFERENCE 1 (bases 1 to 643)
AUTHORS Satou,Y., Shin-I,T., Kohara,Y. and Satoh,N.
TITLE Expressed genes in Clona intestinalis (2002c)
JOURNAL Unpublished
COMMENT Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoh@ascidian.zool.kyoto-u.ac.jp.

FEATURES
source
1..643
/organism="Clona intestinalis"
/mol_type="mRNA"
/db_xref="taxon:7719"
/clone="rc1b062j14"
/tissue_type="whole animal"
/dev_stage="tailbud embryo"
/clone_1ib="Nori Satoh unpublished cDNA library, tailbud
embryo"

BASE COUNT 178 a 145 c 100 g 220 t
ORIGIN

Query Match 0.8%; Score 20; DB 13; Length 643;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2312 ATCATGGCAGAGCAGAAAA 2331
|||||
Db 597 ATCATGGCAGAGCAGAAAA 578
|||||

RESULT 36
BM243771 649 bp mRNA linear EST 09-NOV-2002
LOCUS BM243771 Nori Satoh unpublished cDNA library, tailbud embryo Clona
DEFINITION intestinalis cDNA clone cltb062j14 5', mRNA sequence.
ACCESSION BM243771
VERSION BM243771.1 GI:24823689

KEYWORDS EST.
SOURCE Clona intestinalis
ORGANISM Clona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
Phlebobranchia; Clonidae; Clona.
REFERENCE 1 (bases 1 to 649)
AUTHORS Satou,Y., Shin-I,T., Kohara,Y. and Satoh,N.
TITLE Expressed genes in Clona intestinalis (2002c)
JOURNAL Unpublished
COMMENT Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoh@ascidian.zool.kyoto-u.ac.jp.

FEATURES
source
1..649
/organism="Clona intestinalis"
/mol_type="mRNA"
/db_xref="taxon:7719"
/clone="cltb062j14"
/tissue_type="whole animal"
/dev_stage="tailbud embryo"
/clone_1ib="Nori Satoh unpublished cDNA library, tailbud
embryo"

BASE COUNT 222 a 101 c 144 g 182 t
ORIGIN

Query Match 0.8%; Score 20; DB 13; Length 649;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2312 ATCATGGCAGAGCAGAAAA 2331
|||||
Db 37 ATCATGGCAGAGCAGAAAA 56
|||||

RESULT 37
BG698682 651 bp mRNA linear EST 07-MAY-2001
LOCUS BG698682 602703065F1 NC1_CGAP_Skn3 Homo sapiens cDNA IMAGE:4800280 5',
DEFINITION mRNA sequence.
ACCESSION BG698682
VERSION BG698682.1 GI:13966243
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 651)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: James Cleaver, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA
Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM10691 row: a column: 17
High quality sequence start: 3
High quality sequence stop: 606.
Location/Qualifiers
1..651
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4800280"
/lab_host="DH10B (TI phage-resistant)"

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 658)
 Dunm,D., Aoyagi,A., Barber,M., Beacom,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Rilly,M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausen,A. and Wright,D., Weis,R.
 Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
 Unpublished
 CONTACT: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 1000 Std Error: 0.00
 Plate: 0126 row: P column: 17
 Seq primer: CGTGTGTAACGACGCGCAGT
 Class: plasmid ends
 High quality sequence stop: 658.
 Location/Qualifiers
 1..658
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUCIM0126P17"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /clone_lib="Mouse 10kb plasmid UUCIM library"
 /note="Vector: PMD42nv. Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 225 a 128 c 98 g 207 t
 ORIGIN
 Query Match 0.8%; Score 20; DB 28; Length 658;
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1842 TTCATATATGTTGAGCAT 1861
 |||||
 DB 146 TTCATATATGTTGAGCAT 127

RESULT 41
 BW261677
 LOCUS BW261677 664 bp mRNA linear EST 09-NOV-2002
 DEFINITION BW261677 Nori Satoh unpublished cDNA library, gastrula and neurula
 Ciona intestinalis cDNA clone cign029c12 5', mRNA sequence.
 ACCESSION BW261677
 VERSION BW261677.1 GI:24841595
 KEYWORDS EST.
 SOURCE Ciona intestinalis

ORGANISM Ciona intestinalis
 Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona; Phlebobranchia; Clonidae; Ciona. 1 (bases 1 to 664)
 Satou,Y., Shin-i,T., Kohara,Y. and Satoh,N.
 Expressed genes in Ciona intestinalis (2002c)
 Unpublished
 CONTACT: Nori Satoh
 Department of Zoology
 Kyoto University
 Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
 Tel: 81-75-753-4081
 Fax: 81-75-705-1113
 Email: satoh@ascidian.zool.kyoto-u.ac.jp.
 Location/Qualifiers
 1..664
 /organism="Ciona intestinalis"
 /mol_type="mRNA"
 /db_xref="taxon:7719"
 /clone="cign029c12"
 /tissue_type="whole body"
 /dev_stage="gastrula and neurula"
 /clone_lib="Nori Satoh unpublished cDNA library, gastrula and neurula"

BASE COUNT 220 a 111 c 150 g 183 t
 ORIGIN
 Query Match 0.8%; Score 20; DB 13; Length 664;
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2312 ATCATGCAGAGCAGAAAA 2331
 |||||
 DB 36 ATCATGCAGAGCAGAAAA 55

RESULT 42
 BG911162 668 bp. mRNA linear EST 05-JUN-2001
 LOCUS BG911162
 DEFINITION 60280663p1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4938831
 5', mRNA sequence.
 ACCESSION BG911162
 VERSION BG911162.1 GI:14291638
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo. 1 (bases 1 to 668)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: David N. Louis, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: InCyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LAM10874 row: n column: 16
 High quality sequence stop: 195.
 Location/Qualifiers
 1..668
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4938831"
 /tissue_type="anaplastic oligodendroglioma with 1p/19q loss"
 /lab_host="DH10B (T1 phage-resistant)"
 /clone_lib="NCI_CGAP_Brn67"

/note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.3 kb. Constructed by Life Technologies. Note: this is a NOI_CGAP Library."

BASE COUNT 297 a 178 c 157 g 36 t

Query Match 0.8%; Score 20; DB 12; Length 669;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 749 AACACAGACACACACACA 768
DB 316 AACACAGACACACACACA 335

RESULT 43
LOCUS AJ453442 689 bp mRNA linear EST 22-APR-2002
DEFINITION AJ453442 riken1 Gallus gallus cDNA clone 34g19r1, mRNA sequence.

ACCESSION AJ453442
VERSION AJ453442.1 GI:20263538
KEYWORDS EST;
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
1 (bases 1 to 689)
Buerstedde, J.M.
Gallus gallus bursal lymphocyte EST

AUTHORS Buerstedde, J.M.
TITLE Gallus gallus bursal lymphocyte EST
JOURNAL Unpublished
COMMENT Contact: Buerstedde JM

Cellular Immunology
Heinrich-Pette-Institute
Martinstr. 52, 20251 Hamburg, Germany
Email: URI: http://genetics.hpi.uni-hamburg.de/dt40est.html.
Location/Qualifiers

FEATURES
source 1..689
/organism="Gallus gallus"
/mol_type="mRNA"
/db_xref="taxon:9031"
/clone="34g19r1"
/cell_type="bursal lymphocyte"
/dev_stage="2-3 weeks old"
/clone_lib="riken1"
/note="CB inbred strain"

BASE COUNT 207 a 158 c 142 g 182 t
ORIGIN

Query Match 0.8%; Score 20; DB 9; Length 689;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 751 CACAAGCAACACACACACA 770
DB 197 CACAAGCAACACACACACA 216

RESULT 44
LOCUS BW262397 689 bp mRNA linear EST 09-NOV-2002
DEFINITION BW262397 Nori Satoh unpublished cDNA library, gastrula and neurula Ciona intestinalis cDNA clone cign030m24 5', mRNA sequence.

ACCESSION BW262397
VERSION BW262397.1 GI:24842315
KEYWORDS EST;
SOURCE Ciona intestinalis
ORGANISM Ciona intestinalis

REFERENCE Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona; Phlebobranchia; Clonidae; Ciona.
1 (bases 1 to 689)
Satou, Y., Shin-I, T., Kohara, Y. and Satoh, N.

TITLE Expressed genes in Ciona intestinalis (2002c)
JOURNAL Unpublished
CONTACT: Nori Satoh
COMMENT Department of Zoology
Kyoto University
Sakyo-Ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoh@ascidian.zool.kyoto-u.ac.jp.

FEATURES
source 1..689
/organism="Ciona intestinalis"
/mol_type="mRNA"
/db_xref="taxon:7719"
/clone="cign030m24"
/tissue_type="whole body"
/dev_stage="gastrula and neurula"
/clone_lib="Nori Satoh unpublished cDNA library, gastrula and neurula"

BASE COUNT 230 a 110 c 157 g 192 t
ORIGIN

Query Match 0.8%; Score 20; DB 13; Length 689;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2312 ATCATGGCAGAACGAGAAA 2331
DB 33 ATCATGGCAGAACGAGAAA 52

RESULT 45
LOCUS BF294340 698 bp mRNA linear EST 04-MAY-2001
DEFINITION 004PBX08 Pb cDNA #17, Tommaso Pace, Marta Ponzi, and Clara Frontali
Plasmodium berghei cDNA 5', mRNA sequence.

ACCESSION BF294340
VERSION BF294340.1 GI:13943267
KEYWORDS EST;
SOURCE Plasmodium berghei
ORGANISM Plasmodium berghei

REFERENCE Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
1 (bases 1 to 698)
Carlton, J.M.-R. and Dame, J.B.
The Plasmodium vivax and P. berghei gene sequence tag projects
Parasitol. Today (Regul. Ed.) 16 (10), 409 (2000)
Contact: Dame JB
Department of Pathobiology, College of Veterinary Medicine
University of Florida
2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611-0880, USA
Tel: 352 392 4700
Fax: 352 392 9704
Email: damej@mail.vetmed.ufl.edu
Seq primer: T3.

FEATURES
source 1..698
Location/Qualifiers
/organism="Plasmodium berghei"
/mol_type="mRNA"
/strain="ANKA clone HP (gametocyte producer)"
/db_xref="taxon:5821"
/dev_stage="asynchronous blood stage"
/lab_host="Wistar rats"
/clone_lib="Pb cDNA #17, Tommaso Pace, Marta Ponzi, and Clara Frontali"

/note="Vector: pBluescript II vector DNA, excised from lambda ZAP II.; Site 1: EcoRI; Site 2: XhoI; Total RNA was extracted from asynchronous blood stage forms of the cloned ANKA isolate of P. berghei, grown in Wistar rats to 30% parasitemia and 2-5% gametocytemia. Contaminating host white cells had previously been removed and final host cell contamination estimated to be approximately 5%. PolyA+ RNA was extracted and reverse transcribed using an oligo dt-xhoi primer (lambda ZAP II cDNA cloning kit,

Stratagene). Second strand cDNA was made following the manufacturer's protocol. EcoRI adaptors were ligated to the cDNA, and fragments were ligated into EcoRI/XhoI digested vector."

BASE COUNT 264 a 100 c 102 g 232 t

Query Match 0.8%; Score 20; DB 10; Length 698;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 TTTCCTATATAGATGAAAA 73
|||||
274 TTTCCTATATAGATGAAAA 293

RESULT 46
BBS27573 706 bp mRNA linear EST 25-OCT-2001
LOCUS BBS27573 RIKEN full-length enriched, 15 days embryo head Mus
DEFINITION musculus cDNA clone D93004ID13 3', mRNA sequence.
ACCESSION BBS27573
VERSION BBS27573.2 GI:16444500
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 706)

REFERENCE
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Komno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished
On Jul 28, 2000 this sequence version replaced gi:9579031.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel.: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gs.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Komno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
Wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanabe,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multichannel sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Komno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Komodo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamana,K.I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

FEATURES
source
Location/Qualifiers
1..706

/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="D93004ID13"
/sex="mixed"
/tissue_type="head"
/dev_stage="15 days embryo"
/lab_host="DH10B"
/clone_id="RIKEN full-length enriched, 15 days embryo head"
/note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTVN 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGAGATTCGAGTTATTAATTAATCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLX I"

BASE COUNT 184 a 184 c 159 g 179 t

Query Match 0.8%; Score 20; DB 10; Length 706;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 TTTCCTATATAGATGAAAA 73
|||||
534 TTTCCTATATAGATGAAAA 553

RESULT 47
BM027074 708 bp mRNA linear EST 13-OCT-2002
LOCUS BM027074 Nori Satoh unpublished cDNA library, blood cells Ciona
DEFINITION intestinalis cDNA clone rc1bd085j10 3', mRNA sequence.
ACCESSION BM027074
VERSION BM027074.1 GI:23942881
KEYWORDS EST.
SOURCE Ciona intestinalis
ORGANISM Ciona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona; Phlebobranchia; Cionidae; Ciona.

REFERENCE
AUTHORS Satoh,Y., Satake,M., Azumi,K., Nonaka,M., Shin-I,T., Kohara,Y. and Satoh,N.
Expressed genes in Ciona intestinalis (2002)

TITLE
JOURNAL Unpublished
COMMENT Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto 606-8502, Japan
Tel.: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoh@scidian.zool.kyoto-u.ac.jp.
Location/Qualifiers
1..708
/organism="Ciona intestinalis"
/mol_type="mRNA"

FEATURES
source

```

/db_xref="taxon:7719"
/clone="rcibd085j10"
/tissue_type="blood cells"
/clone_id="Nori Satoh unpublished cDNA library, blood
cells"
BASE COUNT      199 a      159 c      116 g      234 t
ORIGIN
Query Match      0.8%; Score 20; DB 13; Length 708;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy      2312 ATCATGCGAAGACAGAAAA 2331
|||||
703 ATCATGCGAAGACAGAAAA 684

RESULT 48
AV862261      712 bp      mRNA      linear      EST 08-NOV-2001
LOCUS      AV862261/c
DEFINITION      AV862261 Nori Satoh unpublished cDNA library, egg Ciona
intestinalis cDNA clone rcieg32k05 3', mRNA sequence.
ACCESSION      AV862261
VERSION      AV862261.1 GI:16849785
KEYWORDS      EST.
SOURCE      Ciona intestinalis
ORGANISM      Ciona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
Phlebobranchia; Clonidae; Ciona.
REFERENCE      1 (bases 1 to 712)
AUTHORS      Satoh,N., Satou,Y., Kohara,Y. and Shin-I,T.
TITLE      Expressed genes in Ciona intestinalis
JOURNAL      Unpublished
COMMENT      Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoh@acidian.zool.kyoto-u.ac.jp.
Location/Qualifiers
1. 712
/organism="Ciona intestinalis"
/mol_type="mRNA"
/db_xref="taxon:7719"
/clone="rcieg32k05"
/tissue_type="whole animal"
/dev_stage="egg"
/clone_id="Nori Satoh unpublished cDNA library, egg"
BASE COUNT      197 a      156 c      119 g      240 t
ORIGIN
Query Match      0.8%; Score 20; DB 9; Length 712;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy      2312 ATCATGCGAAGACAGAAAA 2331
|||||
546 ATCATGCGAAGACAGAAAA 527

RESULT 49
CA763329      713 bp      mRNA      linear      EST 27-NOV-2002
LOCUS      CA763329
DEFINITION      BR060002B10D10.abi IRRI clones Oryza sativa (indica cultivar-group)
cDNA clone BR060002B10D10.abi similar to unknown protein
[Arabidopsis thaliana], mRNA sequence.
ACCESSION      CA763329
VERSION      CA763329.1 GI:25807368
KEYWORDS      EST.
SOURCE      Oryza sativa (indica cultivar-group)
ORGANISM      Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

```

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
1 (bases 1 to 713)
AUTHORS      Bennett,J., Arumugam,K., Latitte,R., Wen,J. and Bruskewich,R.
TITLE      Rice Microarray
JOURNAL      Unpublished
COMMENT      Contact: Mark Fredricksen
Department of Plant Biology
University of Illinois
1201 W. Gregory Dr., Urbana, IL 61801, USA
Tel: 2172653473
Email: bohnerlab@life.uiuc.edu
International Rice Information System (IRIS);
http://www.iris.irri.org).
Location/Qualifiers
1. 713
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/cultivar="IR64"
/db_xref="taxon:39946"
/clone="BR060002B10D10.abi"
/tissue_type="panicles"
/dev_stage="flowering"
/clone_id="IRRI clones"
/note="Vector: pBluescript II SK+; Water stress was
applied by not watering for 4 consecutive days. Panicles
were collected from control (well watered) and stressed
plants at 2 days before heading, heading, 50% flowering
and 4 days after 50% flowering."
BASE COUNT      200 a      143 c      125 g      243 t
ORIGIN
Query Match      0.8%; Score 20; DB 14; Length 713;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy      384 CAAAAGAGAAATCATCGA 403
|||||
279 CAAAAGAGAAATCATCGA 298

RESULT 50
CNS01FB4      736 bp      DNA      linear      GSS 01-JUN-2001
LOCUS      CNS01FB4
DEFINITION      Anopheles gambiae GSS T7 end of clone 04C02 of NotreDame1 library
from strain PEST of Anopheles gambiae (African malaria mosquito),
genomic survey sequence.
ACCESSION      AL141617
VERSION      AL141617.1 GI:6999735
KEYWORDS      GSS.
SOURCE      Anopheles gambiae (African malaria mosquito)
ORGANISM      Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
Anopheles.
1 (bases 1 to 736)
REFERENCE      Genoscope.
AUTHORS      Direct Submission
TITLE      Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
JOURNAL      BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
2 (bases 1 to 736)
REFERENCE      Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
AUTHORS      Direct Submission
TITLE      Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr.
JOURNAL      Roux, Paris 75015, France
COMMENT      This clone is from an A. gambiae BAC library provided by F.H.
Collins and sequenced by Genoscope in collaboration with the
Laboratory of Biochem. and Biol. Molec. of Insects, Institut
Pasteur.
Location/Qualifiers
1. 736
/organism="Anopheles gambiae"

```

```

/mol_type="genomic DNA"
/strain="PEST"
/db_xref="taxon:7165"
/clone="04C02"
/clone_lib="Notredame1"
/notes="end: 17"

BASE COUNT      175 a      205 c      175 g      177 t      4 others
ORIGIN

Query Match
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2141 CCTGCTGACGACGAACTCCC 2160
Db      358 CCTGCTGACGACGAACTCCC 339

RESULT 51
LOCUS      BM132597 748 bp mRNA linear EST 02-NOV-2002
DEFINITION      BM132597 Nori Satoh unpublished cDNA library, gastrula and neurula
ACCESSION      BM132597
VERSION      BM132597.1 GI:24488996
KEYWORDS      EST.
SOURCE      Ciona intestinalis
ORGANISM      Ciona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
Phlebobranchia; Clonidae; Ciona.
REFERENCE      1 (bases 1 to 748)
AUTHORS      Satou,Y., Shin-I,T., Kohara,Y. and Satoh,N.
TITLE      Expressed genes in Ciona intestinalis (2002c)
JOURNAL      Unpublished
COMMENT      Contact: Nori Satoh
              Department of Zoology
              Kyoto University
              Sakyo-Ku, Kyoto 606-8502, Japan
              Tel: 81-75-753-4081
              Fax: 81-75-705-1113
              Email: satoh@ascidian.zool.kyoto-u.ac.jp.
              Location/Qualifiers
                1..748
                /organism="Ciona intestinalis"
                /mol_type="mRNA"
                /db_xref="taxon:7719"
                /clone="rcign030m24"
                /tissue_type="whole body"
                /dev_stage="gastrula and neurula"
                /clone_lib="Nori Satoh unpublished cDNA library, gastrula
                and neurula"

BASE COUNT      215 a      168 c      121 g      244 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 20; DB 13; Length 748;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2312 ATCATGGCAGAGCAGAAAA 2331
Db      713 ATCATGGCAGAGCAGAAAA 694

RESULT 52
LOCUS      BZ995097 761 bp DNA linear GSS 25-MAR-2003
DEFINITION      BZ995097
ACCESSION      BZ995097
VERSION      BZ995097.1 GI:29238514
KEYWORDS      GSS.
SOURCE      Zea mays
ORGANISM      Zea mays

```

```

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 761)
Whitefaw,C.A., Quackenbush,J., Van Aken,S., Uterback,T., Resnick
A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
Maize Genomics Consortium
Unpublished
Contact: Cathy Whitefaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitefaw@tigr.org
Seq primer: TP
Class: sheared ends.
Location/Qualifiers
  1..761
  /organism="Zea mays"
  /mol_type="genomic DNA"
  /strain="B73"
  /db_xref="taxon:4577"
  /clone="ZMBMTA172L24"
  /clone_lib="ZM 0.6-1.0 KB"
  /note="Vector: pCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
  COT selected genomic DNA library"

BASE COUNT      227 a      114 c      144 g      276 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 20; DB 29; Length 761;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      749 AACACAGCAGACAGCA 768
Db      56 AACACAGCAGACAGCA 37

RESULT 53
LOCUS      BM131850 765 bp mRNA linear EST 02-NOV-2002
DEFINITION      BM131850 Nori Satoh unpublished cDNA library, gastrula and neurula
ACCESSION      BM131850
VERSION      BM131850.1 GI:24488249
KEYWORDS      EST.
SOURCE      Ciona intestinalis
ORGANISM      Ciona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
Phlebobranchia; Clonidae; Ciona.
REFERENCE      1 (bases 1 to 765)
AUTHORS      Satou,Y., Shin-I,T., Kohara,Y. and Satoh,N.
TITLE      Expressed genes in Ciona intestinalis (2002c)
JOURNAL      Unpublished
COMMENT      Contact: Nori Satoh
              Department of Zoology
              Kyoto University
              Sakyo-Ku, Kyoto 606-8502, Japan
              Tel: 81-75-753-4081
              Fax: 81-75-705-1113
              Email: satoh@ascidian.zool.kyoto-u.ac.jp.
              Location/Qualifiers
                1..765
                /organism="Ciona intestinalis"
                /mol_type="mRNA"
                /db_xref="taxon:7719"
                /clone="rcign029c12"
                /tissue_type="whole body"
                /dev_stage="gastrula and neurula"
                /clone_lib="Nori Satoh unpublished cDNA library, gastrula
                and neurula"

BASE COUNT      220 a      171 c      122 g      252 t
ORIGIN

```

Query Match 0.8%; Score 20; DB 13; Length 765;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2312 ATCATGCGAAGACAGAAAA 2331
 |||||
 Db 719 ATCATGCGAAGACAGAAAA 700

RESULT 54
 CDS01MG/c 770 bp DNA linear GSS 14-JUN-2001
 LOCUS Anopheles gambiae GSS SP6 end of clone 21J20 of Notredame1 library
 DEFINITION from strain PEST of Anopheles gambiae (African malaria mosquito),
 genomic survey sequence.

ACCESSION AL150861
 VERSION AL150861.1 GI:7011340
 KEYWORDS GSS.
 SOURCE Anopheles gambiae (African malaria mosquito)
 ORGANISM Anopheles gambiae
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
 Anopheles.
 1 (bases 1 to 770)

REFERENCE
 TITLE Direct Submission
 AUTHORS Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage;
 JOURNAL BP 191 91006 EVRI cedex - FRANCE (E-mail: seqref@genoscope.cns.fr
 - Web: www.genoscope.cns.fr)
 2 (bases 1 to 770)
 REFERENCE
 AUTHORS Roch,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
 TITLE Direct Submission
 JOURNAL Submitted (16-FEB-2000) BMTI, Institut Pasteur, 25, rue du Dr.
 Roux, Paris 75015, France
 This clone is from an A. gambiae BAC library provided by F.H.
 Collins and sequenced by Genoscope in collaboration with the
 Laboratory of Biochem. and Biol. Molec. of Insects, Institut
 Pasteur.

FEATURES
 location/Qualifiers
 1..770
 /organism="Anopheles gambiae"
 /mol_type="genomic DNA"
 /strain="PEST"
 /db_xref="taxon:7165"
 /clone="21J20"
 /clone_lib="Notredame1"
 /note="end : SP6"
 /note="end : SP6"
 BASE COUNT 237 a 152 c 114 g 262 t 5 others
 ORIGIN

Query Match 0.8%; Score 20; DB 29; Length 770;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1594 TTTTGATGAACATGATATTA 1613
 |||||
 Db 440 TTTTGATGAACATGATATTA 421

RESULT 55
 CDS98673 776 bp mRNA linear EST 03-APR-2003
 LOCUS AGNCNCOURT_12972342 NIH_MGC_178 Mus musculus cDNA clone
 IMAGE:3029054 5', mRNA sequence.
 ACCESSION CDS98673
 VERSION CDS98673.1 GI:29516529
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 776)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Dr. Michael Brownstein
 CDNA Library Preparation: Michael Brownstein Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.lnl.gov>
 Plate: NDCM62 row: 1 column: 15
 High quality sequence stop: 419.

FEATURES
 location/Qualifiers
 1..776
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:3029054"
 /lab_host="DH10B (T1-phage-resistant)"
 /clone_lib="NIH_MGC_178"
 /note="Organ: lung and heart; Vector: pDNR-LIB; Site_1:
 SfiI (ggccattatggc); Site_2: SfiI (ggccgctcgcc); CDNA
 made by oligo-dT priming and directionally cloned. 5' and
 3' adaptors were used in cloning as follows:
 5'-AAGCAGTGTATCAACGACAGTGGCCATTACGCCGG-3' and
 5'-ATTCTAGAGCGGAGCGGCGGCGACATG-dt(30)NN-3'. Full-length
 enriched library was constructed using the Clontech
 Creator SMART kit and size-selected to contain the 0.5 kb
 size fraction. Library created in the laboratory of M.
 Brownstein (NIMH, NIH). Note: this is a NIH_MGC Library."

BASE COUNT 208 a 195 c 143 g 230 t
 ORIGIN

Query Match 0.8%; Score 20; DB 14; Length 776;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 942 TCATCCTTACTCTCAATG 961
 |||||
 Db 680 TCATCCTTACTCTCAATG 699

RESULT 56
 BE130716 792 bp mRNA linear EST 20-FEB-2001
 LOCUS L48-922T3 Ice plant Lambda Uni-Zap XR expression library, 48 hours
 DEFINITION NaCl treatment Mesembryanthemum crystallinum cDNA clone L48-922 5',
 mRNA sequence.

ACCESSION BE130716
 VERSION BE130716.1 GI:8578079
 KEYWORDS EST.
 SOURCE Mesembryanthemum crystallinum (common iceplant)
 ORGANISM Mesembryanthemum crystallinum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Caryophyllales; Aizoaceae; Mesembryanthemum.
 1 (bases 1 to 792)

REFERENCE
 Cushman, J.C.
 TITLE An expressed sequence tag database for the common ice plant,
 Mesembryanthemum crystallinum
 JOURNAL Unpublished
 COMMENT Contact: Cushman JC
 Department of Biochemistry
 University of Nevada
 MS200, Reno, NV 89557-0014, USA
 Tel: 775-784-1918
 Fax: 775-784-1650
 Email: jcushman@unr.edu
 PCR Primers
 FORWARD: T7
 BACKWARD: T3

Plate: L48-10 row: B column: 10
 Seq primer: T3
 High quality sequence stop: 350
 POLYA-No.

FEATURES

source

Location/Qualifiers

1..792

/organism="Mesembryanthemum crystallinum"

/mol_type="mRNA"

/db_xref="taxon:3544"

/clone="L48-922"

/issue_type="leaf, 48 h 0.4M NaCl"

/dev_stage="Six week old"

/clone_1ib="Ice plant lambda uni-zap XR expression library"

/48 hours NaCl treatment"

/note="Vector: lambda uni-zap XR, Bluescript SK-, site_1:

EcoRI; site_2: XhoI"

EcoRI; site_2: XhoI"

BASE COUNT

232 a 175 c 154 g 231 t

ORIGIN

Query Match

Best Local Similarity 100.0%; Score 20; DB 10; Length 792;
 Pred. No. 2.9e+02;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1925 TTGCTGCTTGATGATCA 1944

Db

677 TTGCTGCTTGATGATCA 696

RESULT 57

BH374713

795 bp DNA linear GSS 10-DEC-2001

LOCUS

AG-ND-161A24.TR ND-7AM Anopheles gambiae genomic clone AG-ND-161A24

DEFINITION

, genomic survey sequence.

ACCESSION

BH374713.1 GI:17320855

VERSION

GSS.

KEYWORDS

Anopheles gambiae (African malaria mosquito)

SOURCE

Anopheles gambiae

REFERENCE

1 (bases 1 to 795)
 Hong, Y.S., Hogan, J.R., Wang, X., Sarkar, A., Sim, C., Loftus, B.J., Ren, C., Huff, B.R., Carlile, J.L., Black, K., Zhang, H.-B., Gardner, M.J., and Collins, F.H.

AUTHORS

Construction of a BAC library and generation of BAC end

TITLE

Sequence-tagged connectors for genome sequencing of the African

JOURNAL

Mol. Genet. Genomics 268 (6), 720-728 (2003)

MEDLINE

22542063

PUBMED

12655398

COMMENT

Other_GSSs: AG-ND-161A24.TF

CONTACT

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208

Fax: 301 838 3543

Email: b.loftus@tigr.org

This clone is from an A. gambiae BAC library (ND-7AM) provided by

F.H. Collins and sequenced by The Institute for Genomic Research

(TIGR). The BAC library was generated from A. gambiae PEST strain

DNA. All DNA was extracted from newly hatched first instar larvae

to minimize the inclusion of DNA from microorganisms that inhabit

the gut. The DNA is derived from mixed sexes of larvae. The BAC

library was constructed at Texas A&M University BAC Center

University, College Station, Texas 77843-2123, USA using a HindIII

partial digest.

Seg primer: M13 Rev

Classes: BAC ends.

Location/Qualifiers

1..795

FEATURES

/organism="Anopheles gambiae"

/mol_type="genomic DNA"

/strain="PEST"

/db_xref="taxon:7165"

/clone="AG-ND-161A24"

/clone_1ib="ND-7AM"

/note="Vector: pECBAC1; site_1: HindIII"

BASE COUNT

197 a 189 c 207 g 202 t

ORIGIN

Query Match

Best Local Similarity 100.0%; Score 20; DB 28; Length 795;
 Pred. No. 2.9e+02;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

2141 CCTGCTGAGCCAGAGTCCC 2160

Db

204 CCTGCTGAGCCAGAGTCCC 223

RESULT 58

BJS66263/c

809 bp mRNA linear EST 18-DEC-2002

LOCUS

BJS66263 Ipomoea nil mixture of flower and flower bud Ipomoea nil

DEFINITION

cDNA clone jml0122 3', mRNA sequence.

ACCESSION

BJS66263.1 GI:27248083

VERSION

EST.

KEYWORDS

Ipomoea nil (Japanese morning glory)

SOURCE

Ipomoea nil (Japanese morning glory)

ORGANISM

Ipomoea nil (Japanese morning glory)

REFERENCE

1 (bases 1 to 809)
 Hoshino, A., Seki, M., Shin, I.T., Carninci, P., Kamiya, A., Shiraki, T., Nitsakura, E., Shinozaki, K., Hayashizaki, Y., Kohara, Y. and Iida, S.

AUTHORS

ESTs of Japanese morning glory

TITLE

Unpublished

JOURNAL

Center For Genetic Resource Information

COMMENT

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshin@genes.nig.ac.jp.

FEATURES

location/Qualifiers

source

1..809

/organism="Ipomoea nil"

/mol_type="mRNA"

/cultivar="Tokyo-kokoi standard"

/db_xref="taxon:35883"

/clone="jml0122"

/issue_type="mixture of flower and flower bud"

/clone_1ib="Ipomoea nil mixture of flower and flower bud"

BASE COUNT

216 a 186 c 171 g 235 t

ORIGIN

Query Match

Best Local Similarity 100.0%; Score 20; DB 12; Length 809;
 Pred. No. 2.9e+02;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

103 TCCTGATGAGGTAGCAAGC 122

Db

154 TCCTGATGAGGTAGCAAGC 135

RESULT 59

CC159651/c

812 bp DNA linear GSS 29-APR-2003

LOCUS

i903c07.g1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone

DEFINITION

i903c07, genomic survey sequence.

ACCESSION

CC159651.1 GI:30184429

VERSION

GSS.

KEYWORDS

Zea mays

ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 812)
Rabinowicz, P.D., O'Shaughnessy, A.L., Balija, V., Dedhia, N., Kuzemburger, F., King, L., Miller, B., Muller, S., Nascimento, L., Zaveren, T., McCombie, W.R. and Martienssen, R.A.
Genomic shotgun sequences from Zea mays (methyl-filtered)

TITLE Unpublished
JOURNAL
COMMENT Contact: W. Richard McCombie
Lila Annenberg Hazen Genome Sequencing Center
Cold Spring Harbor Laboratory
PO Box 100, Cold Spring Harbor, NY 11724, USA
Tel: 516 367 8884
Fax: 516 367 8874
Email: mcombie@cshl.org
Plate: 1903 row: C column: 07
Seq primer: -21M13UnivRev
Class: shotgun
High quality sequence stop: 812.

FEATURES
source
Location/Qualifiers
1..812
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="B73"
/db_xref="taxon:4577"
/clone="1903C07"
/lab_host="DH5a"
/clone_lib="WGS-ZmaysF (DH5a methyl filtered)"
/note="Organ: immature ears; Site 1: Xba I; Site 2: Xba I; The vector was digested with XbaI and one nucleotide was added by fill in in the recessive 3' end. The genomic DNA was nebulized, end repaired, adaptor ligated and size fractionated using sephadex. The resulting fragments were between 0.8 and 3 kb and were cloned into the vector (X/Y reads in M13mp19, b/g reads in pUC19). The same ligation was transformed into DH5a."

BASE COUNT 240 a 152 c 135 g 285 t
ORIGIN

Query Match 0.8%; Score 20; DB 29; Length 812;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1594 TTTGATGACATGATATTA 1613
|||||
Db 646 TTTGATGACATGATATTA 627

RESULT 60
BH045505 831 bp DNA linear GSS 17-JUL-2001
LOCUS RPCI-24-346H17.TV RPCI-24 Mus musculus genomic clone RPCI-24-346H17
DEFINITION , genomic survey sequence.
ACCESSION BH045505
VERSION BH045505.1 GI:14829902
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muride; Murinae; Mus.
1 (bases 1 to 831)
Zhao, S., Nierman, W., Malek, J., Shatsman, S., Aktinc, B., Levins, M., Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregorgis, E., Russell, D., de Jong, P. and Fraser, C.M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished
Other GSS: RPCI-24-346H17.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200
Fax: 301 838 0208
Email: szhae@igir.org
Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.html>). BAC end page: http://www.tigr.org/tbcbac/bac_ends/mouse/bac_end_intro.html
Plate: 346 row: H column: 17
Seq primer: SP6
Class: BAC ends.

FEATURES
source
Location/Qualifiers
1..831
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-346H17"
/sex="Male"
/cell_type="Spleen/Brain"
/clone_lib="RPCI-24"
/note="Vector: pTRABAC1; Site 1: BamHI; Site 2: BamHI; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTRABAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT 260 a 128 c 180 g 263 t
ORIGIN

Query Match 0.8%; Score 20; DB 28; Length 831;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2370 TAACTAGGAAAAATTAAC 2389
|||||
Db 780 TAACTAGGAAAAATTAAC 799

RESULT 61
CNS07DNK 876 bp DNA linear GSS 08-JUL-2001
LOCUS T3 end of clone BD0A014B06 of library BD0A from strain CBS 94 of
DEFINITION Candida tropicalis, genomic survey sequence.
ACCESSION AL440694
VERSION AL440694.1 GI:12224105
KEYWORDS GSS.
SOURCE Candida tropicalis
ORGANISM Candida tropicalis
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
1 (bases 1 to 876)
Souciet, J.L., Agile, M., Artiguenave, F., Blandin, G., Boletín-Fukuhara, M., Bon, E., Brothier, P., Casaregola, S., de-Montigny, J., Dujon, B., Durieux, P., Lepingle, A., Florente, B., Malpertuy, A., Neuvéglise, C., Ozier-Kalogeropoulos, O., Potier, S., Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M., Wincker, P. and Weissenbach, J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies

TITLE Genomic exploration of the hemiascomycetous yeasts: 16. Candida tropicalis
JOURNAL FEMS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711
PUBMED 11152876
REFERENCE 2 (bases 1 to 876)
Blandin, G., Ozier-Kalogeropoulos, O., Wincker, P., Artiguenave, F. and Dujon, B.
Genomic exploration of the hemiascomycetous yeasts: 16. Candida tropicalis
FEMS Lett. 487 (1), 91-94 (2000)

TITLE
JOURNAL
MEDLINE 20584726
PUBMED 11152891
REFERENCE 3 (bases 1 to 876)
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
Direct Submission

JOURNAL
Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seque@genoscope.cns.fr - Web :
This GSS is part of a random genomic sequencing program of thirteen
Yeast species: *Saccharomyces bayanus* var. *uvarum*, *Saccharomyces*
exiguus, *Saccharomyces servazii*, *Zygosaccharomyces rouxii*,
Saccharomyces kluyveri, *Kluyveromyces thermotolerans*, *Kluyveromyces*
lactis var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia*
angusta, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*,
Candida tropicalis and *Yarrowia lipolytica*. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

FEATURES
source
1..876
/organism="Candida tropicalis"
/mol_type="genomic DNA"
/strain="CBS 94"
/db_xref="taxon:5482"
/clone="BD0A014B06"
/clone_lib="BD0A"
/note="end : T3"
Location/Qualifiers

BASE COUNT
309 a 136 c 152 g 275 t 4 others

Query Match 0.8%; Score 20; DB 29; Length 876;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1521 AAATTGACTACTGAGAC 1540
|||||
108 AAATTGACTACTGAGAC 127

RESULT 62
LOCUS B033498 923 bp mRNA linear EST 24-JAN-2001
DEFINITION 602301896F1 NIH_MGC_87 Homo sapiens CDNA clone IMAGE:4403567 5',
mRNA sequence.
ACCESSION B033498
KEYWORDS EST.
VERSION B033498.1 GI:12425857
SOURCE Homo sapiens
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 923)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM10113 row: 0 column: 24
High quality sequence stop: 649.
Location/Qualifiers

FEATURES
source
1..923
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4403567"
/tissue_type="mammary adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_87"
/note="Organ: breast; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.383 kb. Library enriched for

full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 247 a 235 c 199 g 242 t

Query Match 0.8%; Score 20; DB 10; Length 923;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 926 GGAGTCATTACCTTCAT 945
|||||
Db 738 GGAGTCATTACCTTCAT 757

RESULT 63
LOCUS B0712408 957 bp mRNA linear EST 16-JUL-2002
DEFINITION AGENCOURT_8354094 NIH_MGC_113 Homo sapiens CDNA clone IMAGE:6278639
5', mRNA sequence.
ACCESSION B0712408
VERSION B0712408.1 GI:21851307
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 957)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Mark Watson
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM246 row: 0 column: 24
High quality sequence stop: 605.
Location/Qualifiers

FEATURES
source
1..957
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6278639"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_113"
/note="Organ: spleen; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
BASE COUNT 262 a 224 c 205 g 266 t

Query Match 0.8%; Score 20; DB 13; Length 957;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 758 AACCAACAGCACTTACAG 777
|||||
Db 181 AACCAACAGCACTTACAG 200

RESULT 64
LOCUS AK037932 2185 bp mRNA linear HTC 05-DEC-2002
DEFINITION Mus musculus 16 days neonate thymus cDNA, RIKEN full-length

enriched library, clone: A130064P06 product: interferon regulatory factor 6, full insert sequence.

AK037932
AK037932.1 GI:26086171
HTC: CAP trapper.
SOURCE
ORGANISM
Mus musculus
Mus musculus (house mouse)

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
10349636
2
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Mech. Enzymol. 303, 19-44 (1999)
99279253
PUBMED
10349636
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
MEDLINE
11042159
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Suni, N., Ishi, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanuki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
MEDLINE
11076861
4
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishi, Y., Aizawa, K., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanka, I., Saito, T., Okazaki, Y., Gojohori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Flisbach, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaado, I., Pessle, G., Quackenbush, J., Schriml, L. M., Staudt, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bona, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hochmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombert, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilmberg, J., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohlschki, S. and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
MEDLINE
11217851
5
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2185)

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
10349636
2
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Mech. Enzymol. 303, 19-44 (1999)
99279253
PUBMED
10349636
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
MEDLINE
11042159
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Suni, N., Ishi, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanuki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
MEDLINE
11076861
4
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishi, Y., Aizawa, K., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanka, I., Saito, T., Okazaki, Y., Gojohori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Flisbach, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaado, I., Pessle, G., Quackenbush, J., Schriml, L. M., Staudt, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bona, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hochmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombert, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilmberg, J., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohlschki, S. and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
MEDLINE
11217851
5
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2185)

Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/
Location/Qualifiers
1. 2185
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM:DB:A130064P06"
/db_xref="taxon:10090"
/clone="A130064P06"
/tissue_type="thymus"
/clone_id="RIKEN full-length enriched mouse cDNA library"
/dev_stage="16 days neonate"
1. 2185
/note="interferon regulatory factor 6 (MGI:MGI:1859211, GB|NM_016851, evidence: BLASTN, 99%, match=808)"

misc_feature
BASE COUNT 560 a 509 c 539 g 577 t
ORIGIN
Query Match 0.8%; Score 20; DB 11; Length 2185;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 20; Conservative 0; Mismatches 0; Gaps 0;

Cy 2188 CCAACTCAAGACAGACAG 2207
Db 940 CCAACTCAAGACAGACAG 959

RESULT 65
BH907885
LOCUS 66 bp DNA linear GSS 04-SEP-2002
DEFINITION SALK_044651.46.40.x Arabidopsis thaliana TDNA insertion lines
Arabidopsis thaliana genomic clone SALK_044651.46.40.x, genomic survey sequence.
ACCESSION BH907885
VERSION BH907885.1 GI:22720818
SOURCE GSS.
ORGANISM Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
1 (bases 1 to 66)
1 euroids II; Brassicales; Brassicaceae; Arabidopsie.
REFERENCE
Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadriab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P., Zimmerman, J. and Ecker, J.R.
A sequence-indexed library of insertion mutations in the Arabidopsis Genome
unpublished
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

Tel: 858 453 4100 x1752
 Fax: 858 558 6379
 Email: ecker@sal.k.edu
 This is single pass sequence recovered from the left border of
 TDNA. This sequence lies within an annotated intron of At5g15100.
 Class: TDNA tagged.

FEATURES

source

1. 66
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="Columbia 0"
 /db_xref="taxon:3702"
 /clone="SALK_044651.46.40.x"
 /note="PCR was performed on Arabidopsis thaliana lines
 each of which contains one or more TDNA insertion
 elements. The resultant fragment for each line was
 directly sequenced to determine the genomic sequence at
 the site of insertion. Details of the protocols used can
 be found at http://signal.salk.edu/tdna_protocols.html"

BASE COUNT

32 a 9 c 5 g 20 t

ORIGIN

Query Match 0.8%; Score 19; DB 29; Length 66;
 Best Local Similarity 100.0%; Pred. No. 6e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 393 AATCATGACAAAACA 411

Db 12 AATCATGACAAAACA 30

RESULT 66

LOCUS CC021347 110 bp DNA linear GSS 01-APR-2003
 DEFINITION 3591_1_24_1-A11.2EL_Y_1 3591 - Rescuedu Grid P Zea mays genomic,
 genomic survey sequence.

ACCESSION CC021347

VERSION CC021347.1 GI:29435420

KEYWORDS

SOURCE

ORGANISM

Ze mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoidae; Andropogoneae; Zea.

1 (bases 1 to 110)

Walbot, V.

Maize genomic sequences found using engineered Rescuedu transposon

UNPUBLISHED

Contact: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 725 8221

Email: walbot@stanford.edu

Possible ligation site of ends cut by 2 different endonucleases.
 Reverse complemented post-ligation sequence from source sequence.
 Plate: 3591_1_24_1 row: 9
 Class: transposon-tagged.
 Location/Qualifiers
 1. 110
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /cultivar="mixed Background W23/A188/B73/K55"
 /db_xref="taxon:4577"
 /tissue_type="leaf"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_lib="3591 - Rescuedu Grid P"
 /note="Organ: leaf; Vector: Rescuedu (engineered from
 pBluescript backbone); Site 1: BamHI; Site 2: BglII;
 Rescuedu is a 4.9 Kb, modified maize Mu transposon

FEATURES

source

designed to allow plasmid rescue from total genomic DNA.
 Mu elements insert preferentially into transcription
 units. For more information on Rescuedu, go to the web
 site 'www.zmdb.iastate.edu' and follow the links for
 'Rescuedu.' Grid P was grown at Mokolai in 2002. DNA was
 extracted from leaf strips, double digested using BamHI
 and BglII, and ligated to form circular plasmids. DH10B
 cells were transformed and then screened on LB plates with
 ampicillin."

BASE COUNT

36 a 16 c 19 g 39 t

ORIGIN

Query Match 0.8%; Score 19; DB 29; Length 110;
 Best Local Similarity 100.0%; Pred. No. 6.5e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1212 CATCTGAACGTGTAAAA 1230

Db 80 CATCTGAACGTGTAAAA 62

RESULT 67

LOCUS H74951 136 bp mRNA linear EST 31-OCT-1995
 DEFINITION 539 Random-primed Brassica napus cDNA clone RRM1328, mRNA sequence.
 ACCESSION H74951
 VERSION H74951.1 GI:1048166
 KEYWORDS EST.
 SOURCE Brassica napus (rape)
 ORGANISM Brassica napus
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: Nam, HG

Plant Molecular Genetics Laboratory

Pohang University of Science and Technology

Dept. of Life Science, Sanji Hyojadong, Pohang Kyungbuk 790-784,

Korea

Tel: 825622792111

Fax: 825622792199

Email: nam@vision.postech.ac.kr

Seq primer: M13 reverse

Location/Qualifiers

1. 136
 /organism="Brassica napus"
 /mol_type="mRNA"
 /strain="L. cv Naehan"
 /db_xref="taxon:3708"
 /clone="RRM1328"
 /lab_host="E. coli XL-1 Blue"
 /note="Vector: M13mp18; Site 1: EcoRI; Site 2: EcoRI; ESTs
 were generated from the root cDNA library of Brassica
 napus. The cDNAs were primed with random hexamer and were
 cloned into the M13mp18 vector digested with Eco RI."

BASE COUNT

51 a 19 c 34 g 32 t

ORIGIN

Query Match 0.8%; Score 19; DB 14; Length 136;
 Best Local Similarity 100.0%; Pred. No. 6.7e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 ACCAAGGCTATGTCACTTC 183

Db 96 ACCAAGGCTATGTCACTTC 114

RESULT 68
 BE505229/c
 LOCUS BE505229 175 bp mRNA linear EST 06-AUG-2000
 DEFINITION dc19g09.x1 NICHD XGC L11 Xenopus laevis cDNA clone IMAGE:3397600
 3', mRNA sequence.
 ACCESSION BE505229
 VERSION BE505229.1 GI:9708760
 KEYWORDS EST.
 SOURCE Xenopus laevis (African clawed frog)
 ORGANISM Xenopus laevis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
 Xenopodinae; Xenopus.
 1 (bases 1 to 175)
 NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapb-remail.nih.gov
 Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: Xenopus clones from this library are available
 through the I.M.A.G.E. Consortium/ILNL at: info@image.llnl.gov
 Seq primer: -400P from Gibco
 High quality sequence stop: 161.
 Location/Qualifiers
 1..175
 /organism="Xenopus laevis"
 /mol_type="mRNA"
 /db_xref="taxon:8355"
 /clone="IMAGE:3397600"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.4 Kb. Constructed by Life
 Technologies. Note: This is a Xenopus Gene Collection (XGC
) library."
 BASE COUNT 70 a 25 c 29 g 51 t
 ORIGIN

Query Match 0.8%; Score 19; DB 10; Length 175;
 Best Local Similarity 100.0%; Pred. No. 7e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2285 ACTCTCAATTATGATA 2303
 Db 157 ACTCTCAATTATGATA 139

RESULT 69
 BH867080
 LOCUS BH867080 186 bp DNA linear GSS 05-AUG-2002
 DEFINITION hg33b06.y9 WGS-Zmaysf (JMI07 adapted methyl filtered) Zea mays
 genomic clone hg33b06 5', genomic survey sequence.
 ACCESSION BH867080
 VERSION BH867080.1 GI:22102977
 KEYWORDS GSS.
 SOURCE Zea mays
 ORGANISM Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoidae; Andropogoneae; Zea.
 1 (bases 1 to 186)
 Rabinowicz,P.D., O'Shaughnessy,A.L., Ballja,V., Dedhia,N.,
 Katzenburger,F., King,L., Miller,B., Muller,S., Nascimben,L.,
 Zuberavrn,T., McCombie,W.R. and Martienssen,R.A.
 Genomic shotgun sequences from Zea mays (methyl-filtered)

JOURNAL Unpublished
 COMMENT Contact: W. Richard McCombie
 Lita Annenberg Hazen Genome Sequencing Center
 Cold Spring Harbor Laboratory
 PO Box 100, Cold Spring Harbor, NY 11724, USA
 Tel: 516 367 8894
 Fax: 516 367 8874
 Email: mccombie@cshl.org
 Plate: hg93 row: b column: 06
 Seq primer: -21M13univrev
 Class: shotgun
 High quality sequence stop: 186.
 Location/Qualifiers
 1..186
 /organism="Zea mays"
 /mol_type="Genomic DNA"
 /cultivar="B73"
 /db_xref="taxon:4577"
 /clone="hg93b06"
 /lab_host="JMI07 or DHSa"
 /note="Organ: immature ears; Site 1: Xba I; Site 2: Xba I;
 The vector was digested with XbaI and one nucleotide was
 added by fill in in the recessive 3' end. The genomic DNA
 was nebulized, end repaired, adaptor ligated and size
 fractionated using sephadex. The resulting fragments were
 between 0.8 and 3 kb and were cloned into the vector
 (x/y reads in M13mp19, b/g reads in pUC19). The same
 ligation was transformed in either JMI07 or DHSa."
 BASE COUNT 33 a 65 c 56 g 32 t
 ORIGIN

Query Match 0.8%; Score 19; DB 28; Length 186;
 Best Local Similarity 100.0%; Pred. No. 7.1e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1777 TGGAGATAGTCACAGCT 1795
 Db 41 TGGAGATAGTCACAGCT 59

RESULT 70
 AV236691/c
 LOCUS AV236691 220 bp mRNA linear EST 15-NOV-2001
 DEFINITION AV236691 RIKEN full-length enriched, 10 day neonate skin Mus
 musculus cDNA clone 4732415B08 3', mRNA sequence.
 ACCESSION AV236691
 VERSION AV236691.1 GI:6189204
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 220)
 Kono,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
 Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai
 C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
 Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
 Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata
 Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H.,
 Suzuki,H., Takahashi,F., Tateno,M., Tomioka,N., Tsunoda,Y.,
 Watabiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T.,
 Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
 RIKEN Mouse ESTs (Kono,H., et al. 1999)
 Unpublished
 CONTACT: Yoshinide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center(GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suenho-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216

Email: genome-resgsc.riken.go.jp,
 URL: <http://genome.gsc.riken.go.jp/>
 Sasaki, N., Izawa, M., Watanabe, M., Ozawa, K., Tanaka, T., Yoneda, Y.,
 Matsunaga, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki,
 Y.
 Transcriptional sequencing: A method for DNA sequencing using RNA
 polymerase. *Proc. Natl. Acad. Sci. U.S.A.* 95 (7), 3455-3460 (1998)
 Itoh, M., Kikunishi, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, D.,
 Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,
 Y. and Hayashizaki, Y.
 Automated filtration-based high-throughput plasmid preparation
 system. *Genome Res.* 9 (5), 463-470 (1999)
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning. *Methods Enzymol.* 303,
 19-44 (1999)
 Please visit our web site (<http://genome.gsc.riken.go.jp/>) for
 further details.

FEATURES

source
 Location/Qualifiers
 1..220
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="4732415B08"
 /sex="mixed"
 /tissue_type="skin"
 /dev_stage="10 days neonate"
 /lab_host="DH10B"
 /clone_1ib="RIKEN full-length enriched, 10 day neonate
 skin"
 /note="Site 1: SalI; Site 2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5'
 GAGAGAGAGAGATCCAGACCTCTTTTCTTTTCTTTT 3'], cDNA was
 prepared by using trehalose thermo-activated reverse
 transcriptase and subsequently enriched for full-length by
 cap-trapper. cDNA went through one round of normalization
 to Rot = 10.0 and subtraction to Rot = 100.0. Second
 strand cDNA was prepared with the primer adapter of
 sequence [5' GAGAGAGAGATTCGATTAATTAATTAATCCCCCCCCCCC
 3']. cDNA was cloned into the XhoI and BamHI sites.
 Vector: a modified pBluescript KS(+) after bulk excision
 from Lambda FLC I"

BASE COUNT 67 a 37 c 40 g 76 t
 ORIGIN
 Query Match 0.8%; Score 19; DB 9; Length 220;
 Best Local Similarity 100.0%; Pred. No. 7.3e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 2277 ATAATTGACTCTCAAT 2295
 |||||
 41 ATTAATTGACTCTCAAT 23

RESULT 71
 LOCUS AO948125/c 257 bp DNA linear GSS 27-JUN-2000
 DEFINITION Sheared DNA-47C22. TR Sheared DNA Trypanosoma brucei genomic clone
 AO948125
 VERSION AO948125.1 GI:6771390
 KEYWORDS GSS.
 SOURCE Trypanosoma brucei
 ORGANISM Trypanosoma brucei
 Eukaryota; Euzoenzoa; Kinetoplastida; Trypanosomatidae;
 Trypanosoma.
 1 (bases 1 to 257)
 El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C.,

TITLE
 JOURNAL
 COMMENT
 Other GSSs: Sheared DNA-47C22.TF
 Contact: Najib M. El-Sayed
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: neilsayed@igr.org
 Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
 DNA library constructed at TIGR. Clones will be available for
 distribution through ATCC. Sheared DNA end sequences search page:
<http://www.tigr.org/tdb/mdb/tbdb/>.
 Seq primer: M13-Reverse
 Class: shotgun.

FEATURES

source
 Location/Qualifiers
 1..257
 /organism="Trypanosoma brucei"
 /mol_type="genomic DNA"
 /strain="TREU927/4 GUTat 10.1"
 /db_xref="taxon:5691"
 /clone="Sheared DNA-47C22"
 /clone_1ib="Sheared DNA"
 /note="Vector: pUC18; Site 1: SmaI; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from a cloned population of
 Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
 sheared to give a tight size distribution (approx 2 kb).
 The v + i method used for the library construction is
 described in detail in Smith, H.O. and Venter, J.C.
 (Making small insert libraries for whole genome shotgun
 sequencing projects. In *Genome Sequencing: A Practical
 Approach*, eds. M. Vaudin and B. Barrell, Oxford University
 Press, 1999)."

BASE COUNT 40 a 54 c 58 g 105 t
 ORIGIN
 Query Match 0.8%; Score 19; DB 28; Length 257;
 Best Local Similarity 100.0%; Pred. No. 7.5e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 759 ACAACGCAACCTTAACG 777
 |||||
 194 ACAACGCAACCTTAACG 176

RESULT 72
 LOCUS BB402033 271 bp mRNA linear EST 15-JUL-2000
 DEFINITION BB402033 RIKEN full-length enriched, ES cells Mus musculus cDNA
 clone C330025N03.3 similar to AK001942 Homo sapiens cDNA FLJ11080
 f15 clone PLACB1005181, mRNA sequence.
 BB402033
 |||||
 194 ACAACGCAACCTTAACG 176
 ACCESSION BB402033.1 GI:9221429
 VERSION BB402033.1
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 271)
 Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci,
 P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,
 Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,
 Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,
 Kiyosawa, H., Kojima, Y., Kondo, S., Koyama, S., Kurihara, C., Kusakabe, M.,
 Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
 Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata,
 Y., Shigemoto, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y.,
 Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya

T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamana, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Konno, H., et al.)
Unpublished
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermolabilization and thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Kitanai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.riken.go.jp) for further details.
Location/Qualifiers
1. 271
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="C330025N03"
/cell_type="ES cells"
/lab_host="SOLR"
/lab_host="RIKEN full-length enriched, ES cells"
/note="Site 1: XhoI; Site 2: BamHI. cDNA library was prepared and sequenced in Mouse Genome Encyclopedic Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'
GAGAGAGGAGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 25.0. Second strand cDNA was prepared with the primer adapter of sequence [5'
GAGAGAGATCTTCGAGTAAATTAATTAATCCCCCCCCCC 3']".
BASE COUNT 78 a 61 c 46 g 85 t 1 others
ORIGIN
Query March 0.8%; Score 19; DB 10; Length 271;
Best Local Similarity 100.0%; Pred. No. 7.6e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1420 CTTGATGATGATGACT 1438
|||||
Db 35 CTTGATGATGATGACT 53
RESULT 73
LOCUS CD265572 296 bp mRNA linear EST 23-MAY-2003
DEFINITION pmc011xc05f.287559 smc. Phytophthora sojae grown in synthetic medium
Phytophthora sojae cDNA clone smc011xc05 5, mRNA sequence.
ACCESSION CD265572
VERSION CD265572.1 GI:31053399
KEYWORDS EST.

SOURCE
ORGANISM Phytophthora sojae
Phytophthora sojae
Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
Phytophthora.
REFERENCE
1 (bases 1 to 296)
Tyler, B.M., Judelson, H.S., Gijzen, M., Dean, R.A. and Waugh, M.E.
USDA-JFARS: Phytophthora sojae genes expressed during infection and propagation
Unpublished
Contact: Tyler B
Tyler lab
VBI
1880 Pract Dr., Blacksburg, VA 24061, USA
Tel: 540-231-7318
Email: bmt Tyler@vt.edu
PCR Primers
FORWARD: M13 reverse 17mer at 5' end
BACKWARD: M13 reverse 17mer at 5' end
Plate: 011 row: C column: 05
Seq primer: M13 reverse 17mer at 5' end
High quality sequence stop: 296.
Location/Qualifiers
1. 296
/organism="Phytophthora sojae"
/mol_type="mRNA"
/db_xref="taxon:57593"
/clone="smc011xc05"
/dev_stage="mycelium"
/clone_id="smc: Phytophthora sojae grown in synthetic medium"
/note="vector: pCMV-Sport6.1; Site_1: SalI; Site_2: NotI"
BASE COUNT 83 a 84 c 80 g 49 t
ORIGIN
Query March 0.8%; Score 19; DB 14; Length 296;
Best Local Similarity 100.0%; Pred. No. 7.7e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1805 CGTGAAGGGGGAAC 1823
|||||
Db 261 CGTGAAGGGGGAAC 279
RESULT 74
LOCUS AM880404/c 303 bp mRNA linear EST 23-MAY-2000
DEFINITION QVO-OT0030-100400-188-d04 OT0030 Homo sapiens cDNA, mRNA sequence.
ACCESSION AM880404
VERSION AM880404.1 GI:8042414
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 303)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
JOURNAL MEDLINE
PUBMED 10737800
CONTACT: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=ec2-QV0-OT0030-100
400-188-d04kt3-2000-04-10&ct=1)
Seq primer: puc 18 forward
High quality sequence stop: 230.

FEATURES

source

1. .303
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="OT0030"
/note="Organ: ovary; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORSTS PCR (U.S. Letters Patent Application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT

62 a 65 c 67 g 109 t

ORIGIN

Query Match 0.8%; Score 19; DB 10; Length 303;
Best Local Similarity 100.0%; Pred. No. 7.7e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1224 TTTAAATCTGAACGAA 1242

Db 164 TTTAAATCTGAACGAA 146

RESULT 75

LOCUS BM248898 303 bp mRNA linear EST 07-JUN-2003
DEFINITION K0824E07-3 NIA Mouse 8.5-dpc Whole Embryo cDNA Library (Long) Mus
ACCESSION BM248898
VERSION BM248898
KEYWORDS Musculus cDNA clone NIA:K0824E07 IMAGE:30080694 3', mRNA sequence.

EST. BM248898.2 GI:31503499

SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 303)
Piao, Y., Ko, N.-T., Lim, M.-K. and Ko, M.-S. H.
Construction of long-transcript enriched cDNA libraries from
submicrogram amounts of total RNAs by a universal PCR amplification
method

JOURNAL Genome Res. 11 (9), 1553-1558 (2001)

MEDLINE 21439098

PUBMED 11544199

COMMENT

On Dec 17, 2001 this sequence version replaced gi:17884168.
Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Casell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@igsun.grc.nia.nih.gov
Plate: K0824 row: E column: 07
Seq primer: -21M13 Forward
High quality sequence stop: 303
POLYA=Yes.

FEATURES

source

Location/Qualifiers
1. .303
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="taxon:10090"
/clone="NIA:K0824E07 IMAGE:30080694"
/issue_type="whole embryo including extraembryonic
tissues at 8.5-days postcoitum"

/dev_stage="8.5-days postcoitum"
/lab_host="DH10B"
/clone_lib="NIA Mouse 8.5-dpc Whole Embryo cDNA Library
(Long)"

/note="Vector: pSPORT1 (Invitrogen); Site 1: SalI; Site 2:
NotI; Mouse cDNA project by the Laboratory of Genetics,
National Institute on Aging (NIA), Intramural Research
Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is
a long-transcript enriched cDNA library (Ref. Genome Res.
11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were
extracted from a pool of 13 embryos at 8.5-days
postcoitum. Double-stranded cDNAs were synthesized with an
Oligo (dT) primer (Invitrogen):
5'-GACTAGTCTGATCGCGAGCGCCCTTTT-3' and
9.1 ug of total RNA, treated with T4 DNA polymerase, and
purified by ethanol-precipitation. The cDNAs were ligated
to Lene-linker Lr-Sal4, purified by phenol/chloroform, and
separated from free linkers by Centricon 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Tag polymerase (Takara) with a primer Sal4-S. The
products were purified by phenol/chloroform and Centricon
100. The cDNAs were digested with SalI and NotI enzymes
and cloned into SalI/NotI site of pSPORT1 plasmid vector.
The DH10B E. coli host was transformed with the ligation
mixture by the standard chemical method. The average
insert size is about 2.5 kb. The library was constructed
by Yulan Piao (NIA)."

BASE COUNT

65 a 87 c 67 g 84 t

ORIGIN

Query Match 0.8%; Score 19; DB 12; Length 303;
Best Local Similarity 100.0%; Pred. No. 7.7e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1420 CTTGAATGATGATGACT 1438

Db 89 CTTGAATGATGATGACT 71

RESULT 76

LOCUS BB290438 304 bp mRNA linear EST 09-JUL-2000
DEFINITION BB290438 RIKEN full-length enriched, 2 cells egg Mus musculus cDNA
clone B020045P23 3' similar to J04492 Mouse T-cell receptor 4-1BB
protein mRNA, mRNA sequence.

ACCESSION BB290438.1 GI:8990887

VERSION BB290438

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carlini,
1 (bases 1 to 304)
Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,
Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,
Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,
Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M.,
Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata,
Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suganara, Y.,
Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomioka, N., Toyota,
T., Tanoda, Y., Wachi, H., Watanabe, S., Yamamura, T., Yamazaki, I.,
Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino,
M., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Kono, H., et al.)

COMMENT

Unpublished
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenho-cho, Tsunumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222

Fax: 81-45-503-9216
 Email: genome-res@sc.riken.go.jp,
 URL: http://genome-gsc.riken.go.jp/
 Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
 The most efficient and thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
 Itoh, M., Katsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y.
 Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
 Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

FEATURES
 source
 1..304
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="B020045F23"
 /tissue_type="egg"
 /dev_stage="2 cells"
 /lab_host="DH10B"
 /clone_id="RIKEN full-length enriched, 2 cells egg"
 /note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGATTCGAGTAAATTAATTAATCCCCCCCCCC 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATTCGAGTAAATTAATTAATCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified Bluescript KS(+) after bulk excision from Lambda FLIC I."

BASE COUNT
 ORIGIN
 77 a 98 c 54 g 74 t 1 others

Query Match 0.8%; Score 19; DB 10; Length 304;
 Best Local Similarity 100.0%; Pred. No. 7.8e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1742 CCATCTCCAGCGCAGATG 1760
 |||||
 Db 33 CCATCTCCAGCGCAGATG 51

RESULT 77
 BB367558/c 304 bp mRNA linear EST 12-JUL-2000
 LOCUS BB367558 RIKEN full-length enriched, 16 days embryo head Mus
 DEFINITION Musculus musculus cDNA clone C130038G20 3' similar to L38477 Mus musculus (clone Clabp-1) high mobility group 1 protein (HMG-1) mRNA, mRNA sequence
 BB367558
 BB367558.1 GI:9079387
 EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 304)
 Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,

TITLE
 JOURNAL
 COMMENT
 Iizawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shihagawa, A., Shitaki, T., Sogabe, T., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya, T., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamura, T., Yamakawa, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.
 RIKEN Mouse ESTs (Kono, H., et al.)
 Unpublished
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@sc.riken.go.jp,
 URL: http://genome-gsc.riken.go.jp/
 Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
 The most efficient and thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
 Itoh, M., Katsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y.
 Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
 Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

FEATURES
 source
 1..304
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="C130038G20"
 /sex="mixed"
 /tissue_type="head"
 /dev_stage="16 days embryo"
 /lab_host="DH10B"
 /clone_id="RIKEN full-length enriched, 16 days embryo head"
 /note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGATTCGAGTAAATTAATTAATCCCCCCCCCC 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATTCGAGTAAATTAATTAATCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified Bluescript KS(+) after bulk excision from Lambda FLIC I"

BASE COUNT
 ORIGIN
 67 a 51 c 45 g 141 t

Query Match 0.8%; Score 19; DB 10; Length 304;
 Best Local Similarity 100.0%; Pred. No. 7.8e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1435 GACTAATTAAGAAAATG 1453
 |||||

RESULT 78	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
B2244064	B2244064	CH230-362P21-TV CHORI-230 Segment 2 Rattus norvegicus genomic clone	B2244064	B2244064.1	G1:23904316	GSS.	Rattus norvegicus (Norway rat)
309 bp	DNA	linear	GSS	12-bases	1 to 309	1 (bases 1 to 309)	Rattus norvegicus (Norway rat)
12-bases	1 to 309	1 (bases 1 to 309)	Rattus norvegicus (Norway rat)	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
REFERENCE	AUTHORS						
TITLE	JOURNAL						
COMMENT							
FEATURES	source						
ORIGIN							
BASE COUNT							
Query Match	Best Local Similarity	0.8%; Score 19; DB 29; Length 309;					
Matches	19; Conservative	0; Mismatches 0; Indels 0; Gaps 0;					
OY	452	GATGCTGCTGTTGCCTTGG	470				
Db	122	GATGCTGCTGTTGCCTTGG	140				
RESULT 79	LOCUS	DEFINITION <td>ACCESSION</td> <td>VERSION</td> <td>KEYWORDS</td> <td>SOURCE</td> <td>ORGANISM</td>	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
B2244064	B2244064	CH230-362P21-TV CHORI-230 Segment 2 Rattus norvegicus genomic clone	B2244064	B2244064.1	G1:23904316	GSS.	Rattus norvegicus (Norway rat)
309 bp	DNA	linear	GSS	12-bases	1 to 309	1 (bases 1 to 309)	Rattus norvegicus (Norway rat)
12-bases	1 to 309	1 (bases 1 to 309)	Rattus norvegicus (Norway rat)	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
REFERENCE	AUTHORS						
TITLE	JOURNAL						
COMMENT							
FEATURES	source						
ORIGIN							
BASE COUNT							
Query Match	Best Local Similarity	0.8%; Score 19; DB 29; Length 309;					
Matches	19; Conservative	0; Mismatches 0; Indels 0; Gaps 0;					
OY	452	GATGCTGCTGTTGCCTTGG	470				
Db	122	GATGCTGCTGTTGCCTTGG	140				
RESULT 79	LOCUS	DEFINITION <td>ACCESSION</td> <td>VERSION</td> <td>KEYWORDS</td> <td>SOURCE</td> <td>ORGANISM</td>	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
B2244064	B2244064	CH230-362P21-TV CHORI-230 Segment 2 Rattus norvegicus genomic clone	B2244064	B2244064.1	G1:23904316	GSS.	Rattus norvegicus (Norway rat)
309 bp	DNA	linear	GSS	12-bases	1 to 309	1 (bases 1 to 309)	Rattus norvegicus (Norway rat)
12-bases	1 to 309	1 (bases 1 to 309)	Rattus norvegicus (Norway rat)	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
REFERENCE	AUTHORS						
TITLE	JOURNAL						
COMMENT							
FEATURES	source						
ORIGIN							
BASE COUNT							
Query Match	Best Local Similarity	0.8%; Score 19; DB 29; Length 309;					
Matches	19; Conservative	0; Mismatches 0; Indels 0; Gaps 0;					
OY	452	GATGCTGCTGTTGCCTTGG	470				
Db	122	GATGCTGCTGTTGCCTTGG	140				
RESULT 79	LOCUS	DEFINITION <td>ACCESSION</td> <td>VERSION</td> <td>KEYWORDS</td> <td>SOURCE</td> <td>ORGANISM</td>	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
B2244064	B2244064	CH230-362P21-TV CHORI-230 Segment 2 Rattus norvegicus genomic clone	B2244064	B2244064.1	G1:23904316	GSS.	Rattus norvegicus (Norway rat)
309 bp	DNA	linear	GSS	12-bases	1 to 309	1 (bases 1 to 309)	Rattus norvegicus (Norway rat)
12-bases	1 to 309	1 (bases 1 to 309)	Rattus norvegicus (Norway rat)	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
REFERENCE	AUTHORS						
TITLE	JOURNAL						
COMMENT							
FEATURES	source						
ORIGIN							
BASE COUNT							
Query Match	Best Local Similarity	0.8%; Score 19; DB 29; Length 309;					
Matches	19; Conservative	0; Mismatches 0; Indels 0; Gaps 0;					
OY	452	GATGCTGCTGTTGCCTTGG	470				
Db	122	GATGCTGCTGTTGCCTTGG	14				

ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Mus musculus	1	Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Horii, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shimemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, P., Tomimaga, N., Taya, T., Tsumoto, H., Watanabe, A., Watanabe, S., Yamamura, T., Yamanaoka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.	RIKEN Mouse ESTs (Kono, H., et al.)	Unpublished	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Saitama-cho, Tsukuba, Ibaraki, 305-8575, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-resgsc.riken.go.jp URL: http://genome.resc.riken.go.jp/
Mus musculus	1	Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.	Thermolabile and thermostable enzymes by thermostabilization and thermoactivation of the synthesis of full length cDNA, Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)	RIKEN Mouse ESTs (Kono, H., et al.)	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Saitama-cho, Tsukuba, Ibaraki, 305-8575, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-resgsc.riken.go.jp URL: http://genome.resc.riken.go.jp/
Mus musculus	1	Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.	Thermolabile and thermostable enzymes by thermostabilization and thermoactivation of the synthesis of full length cDNA, Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)	RIKEN Mouse ESTs (Kono, H., et al.)	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Saitama-cho, Tsukuba, Ibaraki, 305-8575, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-resgsc.riken.go.jp URL: http://genome.resc.riken.go.jp/
Mus musculus	1	Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.	Thermolabile and thermostable enzymes by thermostabilization and thermoactivation of the synthesis of full length cDNA, Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)	RIKEN Mouse ESTs (Kono, H., et al.)	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Saitama-cho, Tsukuba, Ibaraki, 305-8575, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-resgsc.riken.go.jp URL: http://genome.resc.riken.go.jp/
Mus musculus	1	Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.	Thermolabile and thermostable enzymes by thermostabilization and thermoactivation of the synthesis of full length cDNA, Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)	RIKEN Mouse ESTs (Kono, H., et al.)	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Saitama-cho, Tsukuba, Ibaraki, 305-8575, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-resgsc.riken.go.jp URL: http://genome.resc.riken.go.jp/
Mus musculus	1	Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.	Thermolabile and thermostable enzymes by thermostabilization and thermoactivation of the synthesis of full length cDNA, Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)	RIKEN Mouse ESTs (Kono, H., et al.)	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Saitama-cho, Tsukuba, Ibaraki, 305-8575, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-resgsc.riken.go.jp URL: http://genome.resc.riken.go.jp/
Mus musculus	1	Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.	Thermolabile and thermostable enzymes by thermostabilization and thermoactivation of the synthesis of full length cDNA, Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)	RIKEN Mouse ESTs (Kono, H., et al.)	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Saitama-cho, Tsukuba, Ibaraki, 305-8575, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-resgsc.riken.go.jp URL: http://genome.resc.riken.go.jp/
Mus musculus	1	Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.	Thermolabile and thermostable enzymes by thermostabilization and thermoactivation of the synthesis of full length cDNA, Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)	RIKEN Mouse ESTs (Kono, H., et al.)	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Saitama-cho, Tsukuba, Ibaraki, 305-8575, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-resgsc.riken.go.jp URL: http://genome.resc.riken.go.jp/
Mus musculus	1	Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.	Thermolabile and thermostable enzymes by thermostabilization and thermoactivation of the synthesis of full length cDNA, Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)	RIKEN Mouse ESTs (Kono, H., et al.)	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Saitama-cho, Tsukuba, Ibaraki, 305-8575, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-resgsc.riken.go.jp URL: http://genome.resc.riken.go.jp/
Mus musculus	1	Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.	Thermolabile and thermostable enzymes by thermostabilization and thermoactivation of the synthesis of full length cDNA, Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)	RIKEN Mouse ESTs (Kono, H., et al.)	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Saitama-cho, Tsukuba, Ibaraki, 305-8575, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-resgsc.riken.go.jp URL: http://genome.resc.riken.go.jp/
Mus musculus	1	Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.	Thermolabile and thermostable enzymes by thermostabilization and thermoactivation of the synthesis of full length cDNA, Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)	RIKEN Mouse ESTs (Kono, H., et al.)	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Saitama-cho, Tsukuba, Ibaraki, 305-8575, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-resgsc.riken.go.jp URL: http://genome.resc.riken.go.jp/
Mus musculus	1	Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.	Thermolabile and thermostable enzymes by thermostabilization and thermoactivation of the synthesis of full length cDNA, Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)	RIKEN Mouse ESTs (Kono, H., et al.)	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Saitama-cho, Tsukuba, Ibaraki, 305-8575, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-resgsc.riken.go.jp URL: http://genome.resc.riken.go.jp/
Mus musculus	1	Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.	Thermolabile and thermostable enzymes by thermost		

Query Match 0.8%; Score 19; DB 10; Length 317;
 Best Local Similarity 100.0%; Pred. No. 7.8e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 904 TAGAGTGTTCAGTCCCA 922
 ||||||||||||||||
 DB 148 TAGAGTGTTCAGTCCCA 166

RESULT 80
 BB463922/c 319 bp mRNA linear EST 22-JUL-2000
 LOCUS BB463922 RIKEN full-length enriched, 12 days embryo spinal ganglion
 DEFINITION Mus musculus cDNA D130084G04 3', mRNA sequence.
 ACCESSION BB463922
 VERSION BB463922.1 GI:9381111
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 319)
 Konomo, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci
 P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,
 Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,
 Izawa, M., Kadoya, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,
 Kiyosawa, H., Kojima, Y., Kondo, S., Koyama, S., Kurihara, C., Kusakabe, M.,
 Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
 Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata
 Y., Shigemoto, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y.,
 Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomioka, N., Toya
 T., Tsunoda, Y., Watanabe, S., Watanabe, S., Yamamura, T., Yamazaki, I.,
 Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino
 M., Muramatsu, M. and Hayashizaki, Y.
 RIKEN Mouse ESTs (Konomo, H., et al.)
 Unpublished
 TITLE RIKEN Mouse ESTs (Konomo, H., et al.)
 JOURNAL Unpublished
 COMMENT Contact: Yoshitake Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Saito-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@rsc.riken.go.jp,
 URL: http://genome-gsc.riken.go.jp/
 Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki
 N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Thermostabilization and thermocatalysis of full length
 cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
 Itoh, M., Katsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
 Tomaru, Y., Carninci, P., Shibata, Y., Oza, Y., Muramatsu, M., Okazaki
 Y. and Hayashizaki, Y.
 Automated filtration-based high-throughput plasmid preparation
 system. Genome Res. 9 (5), 463-470 (1999)
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
 19-44 (1999)
 Please visit our web site (http://genome.etc.riken.go.jp) for
 further details.
 Location/Qualifiers
 1..319
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="D130084G04"
 /tissue_type="spinal ganglion"
 /dev_stage="12 days embryo"
 /lab_host="VDH10B"
 /clone_lib="RIKEN full-length enriched, 12 days embryo
 spinal ganglion"
 /note="Site 1: SalI; Site 2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia"

Query Match 0.8%; Score 19; DB 10; Length 319;
 Best Local Similarity 100.0%; Pred. No. 7.8e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1242 AGTTATCAAAACAGAGAG 1260
 ||||||||||||||||
 DB 163 AGTTATCAAAACAGAGAG 145

RESULT 81
 AV046714/c 324 bp mRNA linear EST 23-NOV-1999
 LOCUS AV046714 Mus musculus adult C57BL/6J testis Mus musculus cDNA clone
 DEFINITION AV046714.2 GI:4866379
 ACCESSION AV046714
 VERSION AV046714.2
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 324)
 Carninci, P., Shibata, K., Oza, Y., Konomo, H., Itoh, M., Aizawa, K.,
 Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara
 A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J.,
 Kikuchi, N., Kojima, Y., Matsuyama, T., Nitsuma, H., Oda, H., Owa, C.,
 Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara
 Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomioka, N., Toya
 T., Tsunoda, Y., Watanabe, S., Watanabe, S., Yamamura, T., Yamazaki, I.,
 Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino
 M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
 RIKEN Mouse ESTs
 RIKEN Mouse ESTs
 Unpublished
 TITLE RIKEN Mouse ESTs
 JOURNAL Unpublished
 COMMENT Contact: Chie Owa
 Genome Science Laboratory
 RIKEN
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
 Tel: 81-298-36-9145
 Fax: 81-298-36-9098
 Email: genome-res@etc.riken.go.jp
 Thermostabilization and thermocatalysis of full length
 cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
 Transcriptional sequencing: A method for DNA sequencing using RNA
 polymerase (Proc. Natl. Acad. Sci. U.S.A. 95 (7):3455-3460 (1998))
 Please visit our web site (http://genome.etc.riken.go.jp) for
 further details.
 Location/Qualifiers
 1..324
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="1700063B04"
 /sex="male"
 /tissue_type="testis"
 /dev_stage="adult"
 /clone_lib="Mus musculus adult C57BL/6J testis"

Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer (5'
 GAGAGAGAGAGCGCGCCCACTCGAGTTTCTTTTCTTTTCTTTT 3'), cDNA was
 transcribed by using trehalose thermo-activated reverse
 transcriptase and subsequently enriched for full-length by
 cap-trapper. Second strand cDNA was prepared with the
 primer adapter of sequence (5'
 GAGAGAGAGATTCGAGTTTATTTATTTATCCCTCCCTCCCTCC 3'). cDNA
 was cleaved with BamHI and XhoI. Vector: a modified
 pBluescript KS(+) after bulk excision from lambda FLC I."

```

BASE COUNT      63 a      92 c      47 g      122 t
ORIGIN
Query Match      0.8%; Score 19; DB 9; Length 324;
Best Local Similarity 100.0%; Pred. No. 7.8e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy      2314 CATGGCAGAACGACGAAAA 2332
Db      174 CATGGCAGAACGACGAAAA 156

RESULT 82
A1547552/c      333 bp mRNA linear EST 22-MAR-1999
LOCUS DEFINITION
A1547552 UI-R-C3-str-e-04-0-UI.s1 UI-R-C3 Rattus norvegicus cDNA clone
A1547552 UI-R-C3-str-e-04-0-UI 3', mRNA sequence.
ACCESSION
A1547552.1 GI:4465040
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 333)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PIRBASE 8889548
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MBRRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: benton-soares@iowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized kidney library cDNA Library Preparation: M.B. Soares Labb
Clone distribution: clones will be available through Research
Genetics (www.resgen.com)
Seq primer: M13 Forward.

FEATURES
Source
Location/Qualifiers
1..333
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-C3-str-e-04-0-UI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/clone_idb="UI-R-C3"
/note="Vector: pRTT3-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C3
library is a subtracted library of a series, ultimately
derived from a mixture of individually tagged normalized
libraries from rat placenta, adult lung, brain, liver,
kidney, heart, spleen, ovary, muscle, and 8, 12 and 18-day
embryos, after a series of subtractions to reduce the
representation of cDNAs from which ESTs had already been
generated. The following serially subtracted libraries
were generated in this process: UI-R-C3, UI-R-C2P, UI-R-C1
, UI-R-C0, UI-R-A1, UI-R-A1. The tag is a string of 3-5
nucleotides present between the Not I site and the
oligo-dT track which allows identification of the library
of origin of a clone within the mixture. The subtracted
library (UI-R-C3) was constructed as follows: PCR amplified

```

BASE COUNT	86 a	83 c	59 g	105 t	
ORIGIN					<p>cDNA inserts from UI-R-C2p clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-C2p library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C3 library. This procedure has been previously described (Bonaldi, Lennon and Soares, Genome Research 6:791-806, 1996)"</p>
Qy	1420	CTTGAATGATGATCGACT	1438		
Db	155	CTTGAATGATGATCGACT	137		
Query Match		0.8%	Score 19;	DB 9;	Length 333;
Best Local Similarity		100.0%;	Pred. No. 7.9e+02;		
Matches	19;	Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0;
RESULT 83					
BH294683					
LOCUS	BH294683		335 bp	DNA	linear
DEFINITION	CH230-122P21.TVB	CHORI-230 Segment 1	Rattus norvegicus	genomic	
ACCESSION	clone CH230-122P21,	genomic survey	sequence.		
VERSION	BH294683				
KEYWORDS	BH294683.1	GI:17207091			
SOURCE	GSF.				
ORGANISM	Rattus norvegicus (Norway rat)				
	Rattus norvegicus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
REFERENCE	1 (bases 1 to 335)				
AUTHORS	Zhao, S., Shelly, J., Shatsman, S., Tsegaye, G., Geer, K., Shvartsbeyn, A., Gebregorgis, E., Overton, L., Russell, D., Chen, D., Riggs, F., de Jong, P. and Fraser, C.M.				
TITLE	Rat BAC End Sequences from Library CHORI-230 EcoRI segment				
JOURNAL	Unpublished				
COMMENT	Other_GSSs: CH230-122P21.TVB				
	Contact: Shaying Zhao				
	Department of Eukaryotic Genomics				
	The Institute for Genomic Research				
	9712 Medical Center Dr., Rockville, MD 20850, USA				
	Tel: 301 838 0200				
	Fax: 301 838 0208				
	Email: szhao@tigr.org				
	Clones are derived from the rat BAC library CHORI-230				
	(http://www.chori.org/bacpac/rat230.htm). For BAC library				
	availability, please contact Pieter de Jong (pdjong@mail.cho.org).				
	Clones may be purchased from BACPAC Resources				
	(http://www.chori.org/bacpac/or_exting_information.htm). BAC end				
	page: http://www.tigr.org/cdb/bac_ends/rat/bac_end_intro.html				
	Plate: 122	row: P	column: 21		
	Seq primer: T7				
	Class: BAC ends.				
FEATURES					
SOURCE	Location/Qualifiers				
	1..335				
	/organism="Rattus norvegicus"				
	/mol_type="genomic DNA"				
	/strain="BN/SENHed/MCW"				
	/db_xref="taxon:10116"				
	/clone="CH230-122P21"				
	/sex="Female"				
	/cell_type="Brain"				
	/clone_id="CHORI-230 Segment 1"				
	/note="Vector: pPARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;				
	CHORI-230 Rat (BN/SENHed/MCW) BAC library produced by				
	Pieter de Jong"				
BASE COUNT	67 a	46 c	81 g	141 t	
ORIGIN					

Query Match 0.8%; Score 19; DB 28; Length 335;
 Best Local Similarity 100.0%; Pred. No. 7.9e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 452 GATGCTGCTGTTGCCTTG 470
 |||||
 Db 71 GATGCTGCTGTTGCCTTG 89

RESULT 84
 AA558478 345 bp mRNA linear EST 08-SEP-1997
 LOCUS nk39d05.s1 NCI CGAP_GC2 Homo sapiens CDNA clone IMAGE:1015881 3',
 DEFINITION mRNA sequence.

ACCESSION AA558478
 VERSION AA558478.1 GI:2328955
 KEYWORDS EST.

ORGANISM Homo sapiens (human)
 SOURCE Homo sapiens

REFERENCE 1 (bases 1 to 345)
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished

JOURNAL COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Stratagene, Inc., David B. Krizman,
 Ph.D.

CDNA Library Arraying: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
 Insert length: 1880 Std Error: 0.00
 Seq primer: -40m3 fwd. ET from Amerham
 High quality sequence stop: 260.
 Location/Qualifiers

FEATURES

source 1..345
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1015881"
 /tissue_type="germ cell tumor"
 /lab_host="SOLR (kanamycin resistant)"
 /clone_lib="NCI CGAP GC2"
 /note="Vector: Bluescript SK-; Site 1: EcoRI; Site 2: XhoI
 ; Cloned unidirectionally. Primer: Oligo dT. Bulk germ
 cell tumor. 5' adaptor sequence: 5' GAAITGGGACGAG 3' 3'
 adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3' Average
 insert size: 1.2 Kb."
 BASE COUNT 89 a 65 c 48 g 143 t
 ORIGIN

Query Match 0.8%; Score 19; DB 9; Length 345;
 Best Local Similarity 100.0%; Pred. No. 7.9e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2298 TGGATACCATAGTATCAT 2216
 |||||
 Db 120 TGGATACCATAGTATCAT 138

RESULT 85
 A0825612/c 360 bp DNA linear GSS 27-AUG-1999
 LOCUS HS_5437_B2_A07_S66E RPCI-11 Human Male BAC library Homo sapiens
 DEFINITION genomic clone Plate=1013 Col=14 Row=B, genomic survey sequence.
 ACCESSION A0825612

VERSION A0825612.1 GI:5791674
 GSS.
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 360)
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 PUBMED 10449764

COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887

Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCI-11. For BAC
 library availability, please contact Pieter de Jong
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
 or from Research Genetics (<http://info@resgen.com>). BAC end Web Server:
<http://www.htsc.washington.edu>
 Plate: 1013 row: B column: 14
 Seq primer: SP6
 Class: BAC ends
 High quality sequence stop: 360.
 Location/Qualifiers

FEATURES

source 1..360
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="Plate=1013 Col=14 Row=B"
 /sex="male"
 /clone_lib="RPCI-11 Human Male BAC library"
 /note="Vector: pBAC3.6; Site 1: EcoRI; Site 2: EcoRI;
 Male blood DNA was isolated from one randomly chosen donor
 and partially digested with a combination of EcoRI and
 EcoRII Methylase. Size selected DNA was cloned into the
 pBAC3.6 vector at EcoRI sites"
 BASE COUNT 68 a 67 c 68 g 144 t 13 others
 ORIGIN

Query Match 0.8%; Score 19; DB 28; Length 360;
 Best Local Similarity 100.0%; Pred. No. 8e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1458 ATGATTATTTGGCATTCCT 1476
 |||||
 Db 340 ATGATTATTTGGCATTCCT 322

RESULT 86
 A0910031 364 bp DNA linear GSS 09-JAN-2001
 LOCUS GSSC6405 Trypanosoma cruzi random genomic library Trypanosoma
 DEFINITION cruzi genomic clone G44B9, genomic survey sequence.
 ACCESSION A0910031 GI:10137177
 VERSION A0910031.3
 KEYWORDS GSS.

ORGANISM Trypanosoma cruzi
 SOURCE Trypanosoma cruzi
 Eukaryota; Eukaryota; Kinetoplastida; Trypanosomatidae;

REFERENCE 1 (bases 1 to 364)
 Agüero,F., Verdun,R., Frasch,A.C.C. and Sanchez,D.O.
 A random sequencing approach for the analysis of the trypanosoma
 cruzi genome: general structure, large gene and repetitive DNA

JOURNAL
MEDLINE
PUBMED
COMMENT

families, and gene discovery
Genome Res. 10 (12), 1996-2005 (2000)
2056489
1116094
On Sep 14, 2000 this sequence version replaced gi:337374.
Contact: Sanchez D.O.
Instituto de Investigaciones Biologicas (Univ. Nac. de Gral
San Martin)
Av. Gral Paz S/N, INTI, Edificio 24, B 1650 KNA, San Martin, Buenos
Aires, Argentina
Tel: (54-11) 4580/7255/7
Fax: (54-11) 4752-9639
Email: dsanchez@lib.unsam.edu.ar
Sequences were basecalled with phred and vector was masked with
crossmatch (see <http://genome.washington.edu>). Sequences were then
trimmed from both ends to remove low quality bases and masked
vector.
Seq primer: T7
Classes: shotgun.
Location/Qualifiers

FEATURES
source

1..364
/organism="Trypanosoma cruzi"
/mol_type="genomic DNA"
/strain="CL-Brener"
/db_xref="taxon:5693"
/clone="G4189"
/cell_type="epimastigote"
/clone_lib="Trypanosoma cruzi random genomic library"
/note="Vector: pBS(-) (Stratagene); T. cruzi DNA was
randomly sheared using a nebulizer and the 1 to 2 Kb range
was gel purified and cloned into the dephosphorylated
HincII site of the vector"

BASE COUNT

100 a 116 c 94 g 54 t

ORIGIN

Query Match 0.8%; Score 19; DB 28; Length 364;
Best Local Similarity 100.0%; Pred. No. 8e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 757 CACACACGACGACCTAAC 775
|||||
94 CACACACGACGACCTAAC 112

Db

RESULT 87
BF376452/c 370 bp mRNA linear EST 24-NOV-2000
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

BF376452
MRI-TN0045-290800-002-e06 TN0045 Homo sapiens cDNA, mRNA sequence.
BF376452.1 GI:11338477
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 370)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922

Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=MR1&ct=MR1-TN0045-290800-002-e06&t3=2000-08-29&t4=1>)
Seq primer: puc 18 forward
High quality sequence stop: 360.
Location/Qualifiers

FEATURES
source

1..370
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="TN0045"
/note="Organ: testis normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
67 a 113 c 80 g 110 t

BASE COUNT

67 a 113 c 80 g 110 t

ORIGIN

Query Match 0.8%; Score 19; DB 10; Length 370;
Best Local Similarity 100.0%; Pred. No. 8e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 AGCAATCATGCTGAGCAA 145
|||||
88 AGCAATCATGCTGAGCAA 70

Db

RESULT 88
BF376477 375 bp mRNA linear EST 24-NOV-2000
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

BF376477
MRI-TN0045-290800-003-h10 TN0045 Homo sapiens cDNA, mRNA sequence.
BF376477
BF376477.1 GI:11338502
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 375)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=MR1&ct=MR1-TN0045-290800-003-h10&t3=2000-08-29&t4=1>)
Seq primer: puc 18 forward
High quality sequence stop: 375.
Location/Qualifiers

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="TN0045"
/notes="Organ: testis normal; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT      115 a      82 c      102 g      75 t      1 others
ORIGIN

Query Match      0.8%; Score 19; DB 10; Length 375;
Best Local Similarity 100.0%; Pred. No. 8e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

127 AGAATCATGCTGAGCAA 145
|||||
302 AGAATCATGCTGAGCAA 320

RESULT 89      377 bp      mRNA      linear      EST 07-JUN-2001
LOCUS      fc60e02.y1 Zebrafish Washu MPIMG EST Danio rerio cDNA clone
DEFINITION      IMAGE:3725786 5', mRNA sequence.
ACCESSION      AI883328
VERSION      AI883328
KEYWORDS      EST.
SOURCE      Danio rerio (zebrafish)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
; Cyprinidae; Danio.
REFERENCE      1 (bases 1 to 377)
AUTHORS      Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
,K., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wyllie,T., Underwood
,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Peterson,B.,
Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
and Wilson,R.
TITLE      Washu zebrafish EST Project 1998
JOURNAL      Unpublished
COMMENT      Other ESTs: fc60e02.x1
Contact: Stephen L. Johnson
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@wustl.edu
cDNA Library Preparation: Matthew Clark. cDNA library Arrayed by:
Matthew Clark. DNA Sequencing by: Washington University Genome
Sequencing Center Clone distribution: Genome Systems, St. Louis,
Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
(web address: www.resgen.com) (email contact: info@resgen.com) and
Resourcenzentrumprimatendbank, Berlin, Germany (web address:
www.rzpd.de)
Seq primer: T3 Et from Amersham
High quality sequence score: 361.
FEATURES
source
1. 377
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="IMAGE:3725786"
/sex="mixed"
/tissue_type="26 somite embryos, adult livers, shield
stage embryos"
/lab_host="XL1-blue MRF"

```

```

/clone_lib="Zebrafish Washu MPIMG EST"
/notes="Vector: pSPORT1; Site 1: NotI; Site 2: SalI; 1st
strand cDNA was primed with a Not I - oligo(dT)15 primer
[5'-GAGTATTGACTCGCATG 110
|||||
298 GAGTATTGACTCGCATG 316

RESULT 90      386 bp      mRNA      linear      EST 16-DEC-2002
LOCUS      BY649148
DEFINITION      BY649148 RIKEN full-length enriched, visual cortex Mus musculus
cDNA clone K530040103 3', mRNA sequence.
ACCESSION      BY649148
VERSION      BY649148
KEYWORDS      EST.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
REFERENCE      1 (bases 1 to 386)
AUTHORS      Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
Nikaido,I., Otsu,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H.,
Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C.,
Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A.,
Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S.,
Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Choehle,C., Corbani
,L.E., Cousins,S., Dalia,E., Dragani,T.A., Fletcher,C.F., Forrest
,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A.,
Gough,J., Grimond,S., Guertincich,S., Hirokawa,N., Jackson,I.J.,
Jarvis,E.D., Kanai,A., Kawai,H., Kawasawa,Y., Kedzierski,R.M.,
King,B.L., Kongaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons
,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki
,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pereira,G.,
Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D.,
Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,I., Ring
,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Sempile,C.A., Setou
,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale
,R.D., Tomita,M., Verardo,R., Wegner,L., Wahlestedt,C., Wang,Y.,
Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yangisawa
,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A.,
Carninci,F., Hayatsu,N., Hirozane-Koshikawa,T., Kono,H., Nakamura
,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,K., Aizawa,K.,
Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii
,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata
,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander
,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE      Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL      Nature 420, 563-573 (2002)

```

MEDLINE
PUBMED
1246881
CONTACT: Yoshihide Hayaishizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenho-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel.: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane,
T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Komno, H., Miyazaki, A.,
Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K.,
Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami,
M., Waki, K., Watanabe, A., Muramatsu, M. and Hayaishizaki, Y. Direct
Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multichipillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by Michela Fagiolini and Takao K. Henrich (
Laboratory for Neuronal Circuit Development Brain Science Institute
RIKEN 2-1 Hirotsawa, Wako-shi, Saitama 351-0198 Japan) whose
assistance we gratefully acknowledge.
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.

FEATURES
source
location/Qualifiers
1..386
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="K530640103"
/tissue_type="visual cortex"
/clone_lib="RIKEN full-length enriched, visual cortex"
BASE COUNT
109 a 86 c 78 g 113 t
ORIGIN

Query Match 0.8%; Score 19; DB 13; Length 386;
Best Local Similarity 100.0%; Pred. No. 8.1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1420 CTTGATGATGATGACT 1438
|||||
Db 36 CTTGATGATGATGACT 54

RESULT 91
LOCUS BF376465 396 bp mRNA linear EST 24-NOV-2000
DEFINITION MRL-TN0045-290800-003-b08 TN0045 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF376465
VERSION BF376465.1 GI:11338490
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 396)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R.,

TITLE
JOURNAL
MEDLINE
PUBMED
20202663
CONTACT: Simpson A.U.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL:
(<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=MR1&t2=MR1-TN0045-290800-003-b08&t3=2000-08-29&t4=1>)
Seq primer: puc 18 forward
High quality sequence start: 16
High quality sequence stop: 375.
location/Qualifiers
1..396
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="TN0045"
/note="Organ: testis normal; Vector: puc18; Site 1: Smal;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT
118 a 86 c 114 g 77 t 1 others
ORIGIN

Query Match 0.8%; Score 19; DB 10; Length 396;
Best Local Similarity 100.0%; Pred. No. 8.1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 127 AGGATCAATGCTGAGCAA 145
|||||
Db 323 AGGATCAATGCTGAGCAA 341

RESULT 92
LOCUS BE002498 401 bp mRNA linear EST 05-JUN-2000
DEFINITION PM3-BN0085-260200-001-e02 BN0085 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE002498
VERSION BE002498.1 GI:8262731
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 401)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R.,
Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F.,
Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bala, G. S., Simpson, D. H.,
Brunstein, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare,
M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and
Simpson, A.U.G.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
10737800
CONTACT: Simpson A.U.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL:
(<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=MR1&t2=MR1-TN0045-290800-003-b08&t3=2000-08-29&t4=1>)
Seq primer: puc 18 forward
High quality sequence start: 16
High quality sequence stop: 375.
location/Qualifiers
1..396
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="TN0045"
/note="Organ: testis normal; Vector: puc18; Site 1: Smal;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT
118 a 86 c 114 g 77 t 1 others
ORIGIN

Query Match 0.8%; Score 19; DB 10; Length 396;
Best Local Similarity 100.0%; Pred. No. 8.1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 127 AGGATCAATGCTGAGCAA 145
|||||
Db 323 AGGATCAATGCTGAGCAA 341

RESULT 92
LOCUS BE002498 401 bp mRNA linear EST 05-JUN-2000
DEFINITION PM3-BN0085-260200-001-e02 BN0085 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE002498
VERSION BE002498.1 GI:8262731
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 401)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R.,
Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F.,
Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bala, G. S., Simpson, D. H.,
Brunstein, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare,
M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and
Simpson, A.U.G.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663

PUBMED 10737800
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?l=et2=PM3-BN0085-260
 200-001-e02&t3=2000-02-26&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 12
 High quality sequence stop: 398.
 Location/Qualifiers
 1. 401
 /organism="Homo sapiens"
 /mol_type="rRNA"
 /db_xref="taxon:9606"
 /dev_stage="Adult"
 /clone_lib="BN0085"
 /note="Organ: breast, normal; Vector: puc18; Site 1: Sma1;
 Site 2: Sma1; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the puc 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."
 BASE COUNT 142 a 68 c 68 g 123 t
 ORIGIN
 Query Match 0.8%; Score 19; DB 10; Length 401;
 Best Local Similarity 100.0%; Pred. No. 8.1e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1550 ATTGCTCAATTAGCTGATA 1568
 ||||||||||||||||||
 Db 224 ATTGCTCAATTAGCTGATA 242
 RESULT 93
 B2941299/c 401 bp DNA linear GSS 13-JUN-2003
 LOCUS B2941299/c
 DEFINITION B2941299 401 bp DNA linear GSS 13-JUN-2003
 GENOMIC SURVEY sequence.
 ACCESSION B2941299
 VERSION B2941299.1 GI:31726682
 KEYWORDS GSS.
 SOURCE Bos taurus (cow)
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.
 1 (bases 1 to 401)
 Larkin, D.M., Everts-van der Wind, A., Rebeiz, M., Schweitzer, P.,
 Bachman, S., Green, S., Campos, E.J., Benson, L.D., Edwards, J., Liu, L.,
 Womack, J.E., de Jong, P.J., and Lewin, H.A.
 A Cattle-Human Comparative Map Built with Cattle BAC-ends and Human
 Genome Sequence
 Unpublished
 Other GSSs: CH240_91P17.TV
 Contact: Harris Lewin
 Department of Animal Sciences
 University of Illinois at Urbana Champaign
 1201 W. Gregory Dr., Urbana, IL 61801, USA
 Tel: 217 333 5998
 Fax: 217 244 5617
 Email: h-lewin@uiuc.edu
 Clones are derived from the bovine BAC library CHORI-240
 (http://www.chori.org/bacpac/bovine240.htm). For BAC library
 availability, please contact Pieter de Jong (pdejong@mail.cho.org).

Clones may be purchased from BACPAC Resources
 (http://www.chori.org/bacpac/ordering/information.htm). This work
 was undertaken as part of the International Bovine BAC Mapping
 Consortium (IBMC) by the University of Illinois at Urbana Champaign
 , USA with funds provided by grant No. AG202-34480-11828 from
 USDA-CSREES and AG99-35205-8534 from USDA/NRI (Livestock Genome
 Sequencing Initiative)
 Plate: 91 row: P column: 17
 Seq primer: SP6
 Class: BAC ends.
 Location/Qualifiers
 1. 401
 /organism="Bos taurus"
 /mol_type="genomic DNA"
 /strain="bred: Hereford"
 /db_xref="taxon:9913"
 /clone="CH240_91P17"
 /sex="Male"
 /cell_type="Blood"
 /clone_lib="CHORI-240"
 /note="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI;
 Hereford bull L1 Domino 99375; CHORI-240 Bovine BAC
 library (Male) produced by Pieter de Jong"
 BASE COUNT 88 a 92 c 73 g 146 t
 ORIGIN
 Query Match 0.8%; Score 19; DB 29; Length 401;
 Best Local Similarity 100.0%; Pred. No. 8.1e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 2163 AAGTAGAGACTGAAAGT 2181
 ||||||||||||||||||
 Db 275 AAGTAGAGACTGAAAGT 257
 RESULT 94
 BF885054 402 bp mRNA linear EST 17-JAN-2001
 LOCUS BF885054
 DEFINITION BF885054 402 bp mRNA linear EST 17-JAN-2001
 ACCESSION BF885054
 VERSION BF885054.1 GI:12275663
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 402)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H.,
 Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 2020263
 10737800
 JOURNAL MEDLINE
 PUBMED 10737800
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?l=MR1&t2=MR1-TN0045-
 031000-013-b07&t3=2000-10-03&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 6
 High quality sequence stop: 402.


```

FEATURES
  source
    Location/Qualifiers
      1.402
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /dev_stage="Adult"
        /clone_1lb="TN0045"
        /note="Organ: testis normal; Vector: puc18; Site 1: SmaI;
          Site 2: SmaI; A mini-library was made by cloning products
          derived from ORESTES PCR (U.S. Letters Patent application
          No. 196,716 - Ludwig Institute for Cancer Research)
          profiles into the pUC 18 vector. Reverse transcription of
          tissue mRNA and cDNA amplification were performed under
          low stringency conditions."
BASE COUNT      118 a      86 c      123 g      75 t
ORIGIN
Query Match      0.8%; Score 19; DB 10; Length 402;
Best Local Similarity 100.0%; Pred. No. 8.1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      127 AGGAATCAATGCTGAGCAA 145
Db      329 AGGAATCAATGCTGAGCAA 347

RESULT 95
BF554284      409 bp      mRNA      linear      EST 12-DEC-2000
LOCUS      UI-R-C1-Kg-a-09-0-UI.r1 UI-R-C1 Rattus norvegicus cDNA clone
ACCESSION      UI-R-C1-Kg-a-09-0-UI 5', mRNA sequence.
VERSION      BF554284
KEYWORDS      BF554284.1 GI:11664014
SOURCE      EST
ORGANISM      Rattus norvegicus (Norway rat)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE      1 (bases 1 to 409)
AUTHORS      Ronaldo, M.F., Lennon, G. and Soares, M.B.
TITLE      Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL      Genome Res. 6 (9), 791-806 (1996)
MEDLINE      97044477
PUBMED      8889548
COMMENT      Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            CDNA Library Preparation: M.B. Soares Lab Clone distribution:
            clones will be available through Research Genetics (www.resgen.com)
            This clone is also available through the I.M.A.G.E. Consortium at
            LNC (info@image.llnl.gov). IMAGE ID= 1763903
            Seq primer: MJ3 Forward.
            Location/Qualifiers
              1.409
                /organism="Rattus norvegicus"
                /mol_type="mRNA"
                /strain="Sprague-Dawley"
                /db_xref="taxon:10116"
                /clone="UI-R-C1-Kg-a-09-0-UI"
                /dev_stage="Adult"
                /lab_host="DH10B (Life Technologies)"
                /clone_1lb="UI-R-C1"
                /note="Vector: pT73D-Pac (Pharmacia) with a modified
                  polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-C1
                  library is a subtracted library derived from the UI-R-Co
                  library, which is a subtracted library derived from the
                  UI-R-A1 and UI-R-E1 libraries. The UI-R-A1 library

```

```

BASE COUNT      120 a      94 c      107 g      88 t
ORIGIN
Query Match      0.8%; Score 19; DB 10; Length 409;
Best Local Similarity 100.0%; Pred. No. 8.2e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1420 CTTGAATGATGAATGCACT 1438
Db      368 CTTGAATGATGAATGCACT 386

RESULT 96
BY605266      412 bp      mRNA      linear      EST 15-DEC-2002
LOCUS      BY605266 RIKEN full-length enriched, visual cortex Mus musculus
ACCESSION      cDNA clone K230038L17.3', mRNA sequence.
VERSION      BY605266
KEYWORDS      BY605266.1 GI:26940448
SOURCE      EST.
ORGANISM      Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 412)
AUTHORS      Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
            Nikaide, I., Oatso, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H.,
            Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C.,
            Gojodori, T., Baldarelli, R., Hill, D.P., Bull, C., Hume, D.A.,
            Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S.,
            Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V., Chothia, C., Corbani,
            L.E., Cousins, S., Dalla, E., Dragani, T.A., Flecher, C.F., Forrest,
            A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A.,
            Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J.,
            Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedierski, R.M.,
            King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lennard, B., Lyons,
            P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki,
            H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pertea, G.,
            Pesole, G., Petrowsky, N., Pillai, R., Pontus, J.U., Qi, D.,
            Ramchandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring,
            B.Z., Ringwald, M., Sandelin, A., Schneider, C., Sempie, C.A., Setou,
            M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale,
            R.D., Tomita, M., Verardo, R., Wagner, L., Wahlstedt, C., Wang, Y.,
            Watanabe, Y., Wells, C., Wilming, L.G., Wymshaw-Boris, A., Yanagisawa,
            M., Yang, I., Yang, L., Yuan, Z., Zavalan, M., Zhu, Y., Zimmer, A.,
            Carinici, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura,
            M., Sakazume, N., Sato, K., Shiraki, T., Maki, K., Kawai, J., Aizawa, K.,
            Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii,
            Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata,
            K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander,
            E.S., Rogers, J., Birney, E. and Hayashizaki, Y.
            Analysis of the mouse transcriptome based on functional annotation

```

TITLE

JOURNAL
MEDLINE
PUBMED
12466851

of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

COMMENT

Contact: Yoshihide Hayashiaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gscc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Aikawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane
, Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Kono, H., Miyazaki, A.,
Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K.,
Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami
, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct
Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in Riken.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by Michela Fagioli and Takao K. Hensch (
Laboratory for Neuronal Circuit Development Brain Science Institute
RIKEN 2-1 Hirosawa, Wako-shi, Saitama 351-0198 Japan) whose
assistance we gratefully acknowledge.
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.

FEATURES

source

Location/Qualifiers
1. .412
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="K230038L17"
/tissue_type="visual cortex"
/clone_lib="RIKEN full-length enriched, visual cortex"
BASE COUNT 118 a 90 c 88 g 115 t 1 others
ORIGIN

Query Match 0.8%; Score 19; DB 13; Length 412;
Best Local Similarity 100.0%; Pred. No. 8.2e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1420 CTTGATGATGATGACT 1438
|||||
Db 61 CTTGATGATGATGACT 79

RESULT 97
AL961018 414 bp mRNA linear EST 27-NOV-2002
LOCUS AL961018 XGC-gastrula Silurana tropicalis cDNA clone TGa103b20 5',
DEFINITION mRNA sequence.
ACCESSION AL961018.1 GI:25784613
VERSION EST.
KEYWORDS Silurana tropicalis (western clawed frog)
SOURCE Silurana tropicalis
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS Taylor, R., Ashurst, J.L., Croning, M.D.R., Zorn, A.M. and Rogers, J.
TITLE Sanger Xenopus tropicalis EST project 2002
JOURNAL Unpublished
COMMENT Contact: Taylor R
Sanger Centre
Hinxton, Cambridgeshire, CB10 1SA, UK

Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: TGa103b20.plcsp6
Sequencing primer: SP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.

FEATURES

source

Location/Qualifiers
1. .414
/organism="Silurana tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="TGa103b20"
/dev_stage="gastrula (stages 10.5-13 mixed)"
/lab_host="Escherichia coli XL1-blue"
/clone_lib="XGC-gastrula"
/note=Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from 5' end of poly A+ RNA from stages
10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated
into pCS107 with EcoRI at the 5' end and NotI at the 3'
end."
BASE COUNT 128 a 76 c 111 g 97 t 2 others
ORIGIN

Query Match 0.8%; Score 19; DB 9; Length 414;
Best Local Similarity 100.0%; Pred. No. 8.2e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1577 ACGTCAGATGCTTACATTT 1595
|||||
Db 273 ACGTCAGATGCTTACATTT 291

RESULT 98
BM248963 415 bp mRNA linear EST 07-JUN-2003
LOCUS BM248963
DEFINITION K0925G07-3 NTA Mouse 8.5-dpc Whole Embryo cDNA Library (Long) Mus
musculus cDNA clone NTA:K0925G07 IMAGE:30080814 3', mRNA sequence.
ACCESSION BM248963
VERSION BM248963.2 GI:31503553
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.
AUTHORS Construction of long-transcript enriched cDNA libraries from
TITLE submicrogram amounts of total RNAs by a universal PCR amplification
method

Genome Res. 11 (9), 1553-1558 (2001)
MEDLINE 11544199
PUBMED 11544199
On Dec 17, 2001 this sequence version replaced gi:17884233.
COMMENT Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@igsun-grc.nia.nih.gov
Plate: K0825 row: G column: 07
Seq primer: -21M13 Forward
High quality sequence stop: 415
POLYA=Yes.

FEATURES
source Location/Qualifiers
1. .415

```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="taxon:10090"
/clone="NIA:K0825G07 IMAGE:30080814"
/tissue_type="whole embryo including extraembryonic
tissues at 8.5-days postcoitum"
/dev_stage="8.5-days postcoitum"
/lab_host="DH10B"
/clone_lib="NIA Mouse 8.5-dpc Whole Embryo cDNA Library
(long)"
/note="Vector: pSPORT1 (Invitrogen); Site 1: SalI; Site 2:
NotI; Mouse cDNA project by the Laboratory of Genetics,
National Institute on Aging (NIA), Intramural Research
Program, NIH (http://jgsun.gsc.nia.nih.gov/cDNA). This is
a long-transcript enriched cDNA library (Ref. Genome Res.
11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were
extracted from a pool of 13 embryos at 8.5-days
postcoitum. Double-stranded cDNAs were synthesized with an
oligo(dT) primer [Invitrogen:
5'-pGACTGTTCTTACATCGCAGCGCCGCTTTTCTTTTCTTTT-3'] from
9.1 ug of total RNA, treated with T4 DNA polymerase, and
purified by ethanol-precipitation. The cDNAs were ligated
to lone-linker L1-SalI, purified by phenol/chloroform, and
separated from free linkers by Centricon 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Taq polymerase (Takara) with a primer SalI-S. The
products were purified by phenol/chloroform and Centricon
100. The cDNAs were digested with SalI and NotI enzymes
and cloned into SalI/NotI site of pSPORT1 plasmid vector.
The DH10B E. coli host was transformed with the ligation
mixture by the about 2.5 kb. The library was constructed
by Yulan Piao (NIA)."
BASE COUNT      116 a      97 c      87 g      115 t
ORIGIN
Query Match
Best Local Similarity 100.0%; Pred. No. 8.2e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1420 CTTGAATGATGATCGACT 1438
      |||||
Db      237 CTTGAATGATGATCGACT 219

RESULT 99
BE505361      417 bp      mRNA      linear      EST 06-AUG-2000
LOCUS      dc23a08.x1 NICHDL_XGC_L11 Xenopus laevis cDNA clone IMAGE:3397910
DEFINITION      3', mRNA sequence.
ACCESSION      BE505361
VERSION      BE505361.1 GI:9708892
KEYWORDS      EST.
SOURCE      Xenopus laevis (African clawed frog)
ORGANISM      Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 417)
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINT)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LINT at: info@image.llnl.gov

```

```

Seq primer: -40UP from Gibco
High quality sequence scop: 414.
FEATURES
source
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="IMAGE:3397910"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NICHDL_XGC_L11"
/note="Organ: liver; Vector: pCMV-SPORT6, Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.4 kb. Constructed by Life
Technologies. Note: This is a Xenopus Gene Collection (XGC
) library."
BASE COUNT      132 a      75 c      75 g      134 t      1 others
ORIGIN
Query Match
Best Local Similarity 100.0%; Pred. No. 8.2e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      2285 ACTCTCAATTATGATGATA 2303
      |||||
Db      142 ACTCTCAATTATGATGATA 124

RESULT 100
BY457007      418 bp      mRNA      linear      EST 13-DEC-2002
LOCUS      BY457007 RIKEN full-length enriched, adult male testis Mus musculus
DEFINITION      cDNA clone K630095C12 3', mRNA sequence.
ACCESSION      BY457007
VERSION      BY457007.1 GI:26755150
KEYWORDS      EST.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 418)
REFERENCE
AUTHORS      Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
Nikaido,I., Otsu,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H.,
Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C.,
Gojibori,T., Baldarelli,R., Hill,D.P., Bull,C., Hume,D.A.,
Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S.,
Beisel,K.W., Blake,U.A., Bradt,D., Brusic,V., Chotila,C., Corbani
,L.E., Cousins,S., Dalla,E., Dragani,T.A., Flecher,C.F., Forrest
,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Giesi,C., Godzik,A.,
Gough,J., Grimond,S., Gustincich,S., Hirokawa,N., Jackson,I.J.,
Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedziarski,R.M.,
King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons
,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki
,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G.,
Pesole,G., Petrowsky,N., Pillai,R., Pontius,J.U., Qi,D.,
Ramchandran,S., Ravasi,T., Reed,J.C., Reed,J.U., Reid,J., Ring
,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Sempke,C.A., Setou
,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale
,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlstedt,C., Wang,Y.,
Watanabe,Y., Wells,C., Wilting,L.G., Wynshaw-Boris,A., Yanagisawa
,M., Yang,I., Yang,L., Yuan,Z., Zavalan,M., Zhu,Y., Zimmer,A.,
Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Kono,H., Nakamura
,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
Atakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imclati,K., Ishii
,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata
,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander
,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
JOURNAL
MEDLINE
PUBMED
22354683
12466851
TITLE
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
COMMENT

```

Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@sc.riken.go.jp,
 URL: <http://genome.gsc.riken.go.jp/>

Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M., and Hayashizaki, Y. Direct Submission

Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

FEATURES

source

Location/Qualifiers

1. 418
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="K630095C12"
 /sex="male"
 /tissue_type="testis"
 /dev_stage="adult"
 /lab_host="SOLR"
 /clone_lib="RIKEN full-length enriched, adult male testis"
 /note="Site 1: XhoI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5',

GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGAGAGCGCGCGCATTAATTCGAGTTAATAATTAATCCCGCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites."

BASE COUNT

120 a 87 c 88 g 123 t

ORIGIN

Query Match 0.8%; Score 19; DB 13; Length 418;
 Best Local Similarity 100.0%; Pred. No. 8.2e+02;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1816 GGAAGAAACGAATTCACATC 1834

Db 74 GGAAGAAACGAATTCACATC 56

Search completed: November 14, 2003, 07:49:18
 Job time : 5033 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 03:35:31 ; Search time 147 Seconds
(without alignments)
7173.227 Million cell updates/sec

Title: US-09-765-271-55

Perfect score: 2389

Sequence: 1 TTCTTACGAGTTGGAGCTGT.....TAAGTACGAAAAAATAAAC 2389

Scoring table: OLIGO_NUC
Gapop 60.0, Gapext 60.0

Searched: 569978 seqs, 220691566 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 200 summaries

Database: Issued Patents NA:*

1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
4: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
5: /cgn2_6/prodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfillseq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2389	100.0	2389	3	US-08-961-083-55
2	2389	100.0	2389	4	US-09-536-784-55
3	2389	100.0	2451	4	US-09-468-656A-9
4	2338	97.9	8195	4	US-08-961-527-94
5	53	2.2	2290	3	US-08-961-083-65
6	53	2.2	2290	4	US-09-536-784-65
7	53	2.2	2359	4	US-08-961-527-243
8	53	2.2	2531	4	US-09-468-656A-5
9	53	2.2	2531	4	US-09-468-656A-11
10	44	1.8	973	4	US-08-961-527-355
11	30	1.3	40	3	US-08-961-083-282
12	30	1.3	40	4	US-09-536-784-282
13	30	1.3	40	4	US-09-468-656A-3
14	27	1.1	36	4	US-09-468-656A-1
15	27	1.1	37	3	US-08-961-083-281
16	27	1.1	37	4	US-09-536-784-281
17	26	1.1	35	4	US-09-468-656A-2
18	24	1.0	1342	3	US-08-961-083-181
19	24	1.0	1342	4	US-09-536-784-181
20	24	1.0	1455	4	US-09-468-656A-7
21	19	0.8	6867	4	US-08-961-527-192
22	19	0.8	682	3	US-08-998-416-895
23	19	0.8	7493	1	US-08-212-133A-7
24	19	0.8	7493	1	US-08-474-503-5
25	19	0.8	7493	2	US-08-670-707A-5
26	19	0.8	7493	3	US-09-37-601-5
27	19	0.8	7493	4	US-09-315-179-5

28	19	0.8	7493	5	PCT-US94-13200-5	Sequence 5, Appl
29	18	0.8	499	4	US-09-641-638-52	Sequence 52, Appl
30	18	0.8	499	4	US-09-641-638-572	Sequence 572, Appl
31	18	0.8	659	4	US-09-221-017B-390	Sequence 390, Appl
32	18	0.8	987	4	US-09-107-532A-2361	Sequence 2361, Appl
33	18	0.8	1161	4	US-09-107-532A-1110	Sequence 1110, Appl
34	18	0.8	1664976	4	US-08-916-421B-1	Sequence 1, Appl
35	17	0.7	20	2	US-08-448-386A-10	Sequence 10, Appl
36	17	0.7	20	3	US-08-816-426-10	Sequence 10, Appl
37	17	0.7	20	5	PCT-US93-12161-10	Sequence 10, Appl
38	17	0.7	24	3	US-08-981-462-56	Sequence 56, Appl
39	17	0.7	238	4	US-09-491-356C-15	Sequence 15, Appl
40	17	0.7	253	4	US-09-491-356C-13	Sequence 13, Appl
41	17	0.7	265	4	US-09-491-356C-14	Sequence 14, Appl
42	17	0.7	265	4	US-09-491-356C-16	Sequence 16, Appl
43	17	0.7	265	4	US-09-491-356C-17	Sequence 17, Appl
44	17	0.7	265	4	US-09-491-356C-18	Sequence 18, Appl
45	17	0.7	321	1	US-08-175-388-2	Sequence 2, Appl
46	17	0.7	321	2	US-08-779-620-2	Sequence 2, Appl
47	17	0.7	321	2	US-08-818-726-2	Sequence 2, Appl
48	17	0.7	322	3	US-08-961-083-149	Sequence 149, Appl
49	17	0.7	322	4	US-09-536-784-149	Sequence 149, Appl
50	17	0.7	470	4	US-09-056-556-216	Sequence 216, Appl
51	17	0.7	470	4	US-09-072-596-211	Sequence 211, Appl
52	17	0.7	633	4	US-09-134-001C-1953	Sequence 1953, Appl
53	17	0.7	658	4	US-09-221-017B-1007	Sequence 1007, Appl
54	17	0.7	687	4	US-09-328-352-1650	Sequence 1650, Appl
55	17	0.7	768	3	US-08-981-462-1	Sequence 1, Appl
56	17	0.7	768	4	US-09-470-191-60	Sequence 60, Appl
57	17	0.7	762	4	US-09-071-035-373	Sequence 373, Appl
58	17	0.7	876	4	US-09-501-558-1	Sequence 1, Appl
59	17	0.7	882	4	US-09-501-558-3	Sequence 3, Appl
60	17	0.7	909	4	US-09-252-991A-1381	Sequence 1381, Appl
61	17	0.7	909	4	US-09-328-352-3547	Sequence 3547, Appl
62	17	0.7	975	4	US-09-107-532A-2698	Sequence 2698, Appl
63	17	0.7	1077	3	US-08-839-711-6	Sequence 6, Appl
64	17	0.7	1080	4	US-09-328-352-3376	Sequence 3376, Appl
65	17	0.7	1092	4	US-09-252-991A-1287	Sequence 1287, Appl
66	17	0.7	1230	4	US-09-252-991A-1184	Sequence 1184, Appl
67	17	0.7	1543	4	US-09-359-301A-9	Sequence 9, Appl
68	17	0.7	1578	4	US-09-107-532A-2089	Sequence 2089, Appl
69	17	0.7	1623	4	US-09-613-303-3	Sequence 3, Appl
70	17	0.7	1654	4	US-08-487-596-3	Sequence 3, Appl
71	17	0.7	1756	2	US-08-466-589-3	Sequence 3, Appl
72	17	0.7	1756	2	US-08-700-638-3	Sequence 3, Appl
73	17	0.7	1756	3	US-08-467-578-3	Sequence 3, Appl
74	17	0.7	1756	4	US-09-217-345-3	Sequence 3, Appl
75	17	0.7	1917	3	US-08-755-587-1	Sequence 1, Appl
76	17	0.7	1920	4	US-09-613-303-16	Sequence 16, Appl
77	17	0.7	1947	4	US-09-613-303-28	Sequence 28, Appl
78	17	0.7	2018	3	US-09-370-807-11	Sequence 11, Appl
79	17	0.7	2018	4	US-09-921-259-11	Sequence 11, Appl
80	17	0.7	2255	3	US-08-714-918-105	Sequence 105, Appl
81	17	0.7	2255	3	US-09-265-315-105	Sequence 105, Appl
82	17	0.7	2255	3	US-09-265-315-105	Sequence 105, Appl
83	17	0.7	2255	3	US-09-266-417-105	Sequence 105, Appl
84	17	0.7	2355	4	US-08-913-155-12	Sequence 12, Appl
85	17	0.7	2381	2	US-08-318-826A-9	Sequence 9, Appl
86	17	0.7	2400	6	5215909-13	Patent No. 5215909
87	17	0.7	2416	2	US-08-318-826A-8	Sequence 8, Appl
88	17	0.7	2416	3	US-09-334-489-2	Sequence 2, Appl
89	17	0.7	2416	3	US-09-334-489-2	Sequence 2, Appl
90	17	0.7	2442	1	US-08-175-388-3	Sequence 3, Appl
91	17	0.7	2442	2	US-08-779-620-3	Sequence 3, Appl
92	17	0.7	2842	2	US-08-818-726-3	Sequence 3, Appl
93	17	0.7	2842	4	US-09-613-303-20	Sequence 20, Appl
94	17	0.7	3882	4	US-09-484-970B-68	Sequence 68, Appl
95	17	0.7	4380	4	US-08-955-565A-3	Sequence 3, Appl
96	17	0.7	4643	4	US-09-436-874-1	Sequence 1, Appl
97	17	0.7	6994	4	US-09-491-356C-2	Sequence 2, Appl
98	17	0.7	6999	1	US-08-276-594A-1	Sequence 1, Appl
99	17	0.7	7056	1	US-08-121-202-1	Sequence 1, Appl
100	17	0.7	8241	6	5171844-1	Patent No. 5171844

[illegible][illegible]

Query Match 100.0%; Score 2389; DB 3; Length 2389;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTTAAGAGTTGGGAGCTGTATCAAGCTTAAGACGGTTAAGAAAATTAATCGTGTCTTA 60
 DB 1 TTCTTAAGAGTTGGGAGCTGTATCAAGCTTAAGACGGTTAAGAAAATTAATCGTGTCTTA 60

QY 61 TTTAGATGGAAGAAACGAGCAAGCAAAACGAGAAATTTGACTCCGTGAGAGTTAGCA 120
 DB 61 TTTAGATGGAAGAAACGAGCAAGCAAAACGAGAAATTTGACTCCGTGAGAGTTAGCA 120

QY 121 GGGTGAAGAAATCAATAGCTGAGCAAACTGTCATCAAGATAACAGACCAAGCTATGAC 180
 DB 121 GGGTGAAGAAATCAATAGCTGAGCAAACTGTCATCAAGATAACAGACCAAGCTATGAC 180

QY 181 TTTCACATGCGCAGCACTATCATTTATTAATGTAAGGTTCTTATGACGCTATCATCG 240
 DB 181 TTTCACATGCGCAGCACTATCATTTATTAATGTAAGGTTCTTATGACGCTATCATCG 240

QY 241 TGAAGAAATTAATCAATGAAGATCCAACTTAAGCTTAAGAAAGATGAGGATATGTTAATA 300
 DB 241 TGAAGAAATTAATCAATGAAGATCCAACTTAAGCTTAAGAAAGATGAGGATATGTTAATA 300

QY 301 GGTCAAGGAGTGAATATGTTATCAAGGTAGTGAATACTAATGTTTACCTTAAGATGC 360
 DB 301 GGTCAAGGAGTGAATATGTTATCAAGGTAGTGAATACTAATGTTTACCTTAAGATGC 360

QY 361 TGCCCAACGCGGATTAACGTCGTAACAAAGAGAAATCAATCGACAAAACAAAGAGCATAG 420
 DB 361 TGCCCAACGCGGATTAACGTCGTAACAAAGAGAAATCAATCGACAAAACAAAGAGCATAG 420

QY 421 TCAACATCGTGAAGGTGGAACCTCAAGAAACGATGCTGTGCTTGCTTGACGACGTTCCGA 480
 DB 421 TCAACATCGTGAAGGTGGAACCTCAAGAAACGATGCTGTGCTTGCTTGACGACGTTCCGA 480

QY 481 AGGAGCCTATCACTACAGATGATGTTATATCTTTAATGCTTCGATATCATAGAGATAC 540
 DB 481 AGGAGCCTATCACTACAGATGATGTTATATCTTTAATGCTTCGATATCATAGAGATAC 540

QY 541 TGGTGAATGCTTAATATGCTTCCTATGAGATCAATTAACAATTAATCTTAAGATAGTT 600
 DB 541 TGGTGAATGCTTAATATGCTTCCTATGAGATCAATTAACAATTAATCTTAAGATAGTT 600

QY 601 ATCAGCTACGAGTTGGCTGCTGCAAGAACCTTCCTATCTGGTCGAGAAATCTGTCAAA 660
 DB 601 ATCAGCTACGAGTTGGCTGCTGCAAGAACCTTCCTATCTGGTCGAGAAATCTGTCAAA 660

QY 661 TTCAAGAACTTATCGCCGCAAAATAGCGATAACATTTGAAGAAACAACTGGGTACCTTC 720
 DB 661 TTCAAGAACTTATCGCCGCAAAATAGCGATAACATTTGAAGAAACAACTGGGTACCTTC 720

QY 721 TGTAAAGCAATCCAGGAATCAAAATTAATAACAAGCAACACAGACCACTTAACAGTCA 780
 DB 721 TGTAAAGCAATCCAGGAATCAAAATTAATAACAAGCAACACAGACCACTTAACAGTCA 780

QY 781 AGCAAGTCAAAGTATGACATTTGATGCTCTTGAACACGCTCTTAACAACTGCTTTGAG 840
 DB 781 AGCAAGTCAAAGTATGACATTTGATGCTCTTGAACACGCTCTTAACAACTGCTTTGAG 840

QY 841 TCAAGCAGCATGTAGATCTGATGCGCTTGTCTTGTATCCAGCAACAATACAAGGTGAAC 900
 DB 841 TCAAGCAGCATGTAGATCTGATGCGCTTGTCTTGTATCCAGCAACAATACAAGGTGAAC 900

QY 901 AGCTAGAGGTGTTGACGTCGACACGAGATCAATTAACAATCACTTACCTTAAGTCAAT 960
 DB 901 AGCTAGAGGTGTTGACGTCGACACGAGATCAATTAACAATCACTTACCTTAAGTCAAT 960

QY 961 GTCTGAATTTGGAAGAAAGATGCTGTATTAATCCCTTCTGTTATGCTTCAACCATTTG 1020
 DB 961 GTCTGAATTTGGAAGAAAGATGCTGTATTAATCCCTTCTGTTATGCTTCAACCATTTG 1020

QY 1021 GGTACAGATTCAGGCCAGAAACAAGTCCACACGCACTCCGGAACCTAGTCCAGG 1080
 DB 1021 GGTACAGATTCAGGCCAGAAACAAGTCCACACGCACTCCGGAACCTAGTCCAGG 1080

QY 1081 CCCGCAACCTGCAACCAATCTTAAATAGACTCAAAATCTTCTTGTGTTAGTCACTGGT 1140
 DB 1081 CCCGCAACCTGCAACCAATCTTAAATAGACTCAAAATCTTCTTGTGTTAGTCACTGGT 1140

QY 1141 ACGAAAGTTGGGAGAGATATGTAATTCGAAGAAAGGGACCTCTCGTTATGCTTTGC 1200
 DB 1141 ACGAAAGTTGGGAGAGATATGTAATTCGAAGAAAGGGACCTCTCGTTATGCTTTGC 1200

QY 1201 GAAAGATTTACCATCTGAACCTGTTAAATCTTGAAAGCAAGTATCAAAACAAGAG 1260
 DB 1201 GAAAGATTTACCATCTGAACCTGTTAAATCTTGAAAGCAAGTATCAAAACAAGAG 1260

QY 1261 TGTTCACACACTTAACTGCTAAAAAAGAAATGTGCTCTCTGTCGACCAAGAAATTTTA 1320
 DB 1261 TGTTCACACACTTAACTGCTAAAAAAGAAATGTGCTCTCTCTGTCGACCAAGAAATTTTA 1320

QY 1321 TGATTAAGCATTAATCTGTTAATCTGAGGCTCATTAAGCTTGTGNAATAAGGCTCG 1380
 DB 1321 TGATTAAGCATTAATCTGTTAATCTGAGGCTCATTAAGCTTGTGNAATAAGGCTCG 1380

QY 1381 TAATCTGATTTCCAAAGCCTTAGACAATAATTATGAACGCTTAATGAATGAACTGA 1440
 DB 1381 TAATCTGATTTCCAAAGCCTTAGACAATAATTATGAACGCTTAATGAATGAACTGA 1440

QY 1441 TAAAGAAAAATGTTAGATGATTTATTTGCAATCTCTAGACCAATTAACCATCCAGAGG 1500
 DB 1441 TAAAGAAAAATGTTAGATGATTTATTTGCAATCTCTAGACCAATTAACCATCCAGAGG 1500

QY 1501 ACTTGCAAAACCAATTTCTCAAAATGAGTATACGAACCAAGCAAGTTCGATTCGAAT 1560
 DB 1501 ACTTGCAAAACCAATTTCTCAAAATGAGTATACGAACCAAGCAAGTTCGATTCGAAT 1560

QY 1561 AGCTGATAAGTATACAGAGTCAAGATGTTTACATTTTATGATGAACATATATAGTGA 1620
 DB 1561 AGCTGATAAGTATACAGAGTCAAGATGTTTACATTTTATGATGAACATATATAGTGA 1620

QY 1621 TGAAGAGATGCAATGTAAACGCTCATATGCGCATATGCGCATGAGATTTGAAAAAGATAG 1680
 DB 1621 TGAAGAGATGCAATGTAAACGCTCATATGCGCATATGCGCATGAGATTTGAAAAAGATAG 1680

QY 1681 CCTTCTGATTAAGAAAAAGTTGCACTCAAGCTTAATTAAGAAAAAGATATCTTACC 1740
 DB 1681 CCTTCTGATTAAGAAAAAGTTGCACTCAAGCTTAATTAAGAAAAAGATATCTTACC 1740

QY 1741 TCCATCTCCAGACGAGATGTTAAAGCAAAATCCAACTGGAGATGAGCAGAGCTATTTA 1800
 DB 1741 TCCATCTCCAGACGAGATGTTAAAGCAAAATCCAACTGGAGATGAGCAGAGCTATTTA 1800

QY 1801 CAATGCTGTAAGAGGAAAAAAGCAATTCCTGTTGAGCTTCATATATGTTGAGCA 1860
 DB 1801 CAATGCTGTAAGAGGAAAAAAGCAATTCCTGTTGAGCTTCATATATGTTGAGCA 1860

QY 1861 TACAGTGAAGTTAAAAACGTAATTTGATTAATCTCATATAGGATCATTAACAATAAT 1920
 DB 1861 TACAGTGAAGTTAAAAACGTAATTTGATTAATCTCATATAGGATCATTAACAATAAT 1920

QY 1921 TAAATTTGCTGTTGATGATCACACATACAAAGCTCCAAATGCGTATATACCTTGAAGA 1980
 DB 1921 TAAATTTGCTGTTGATGATCACACATACAAAGCTCCAAATGCGTATATACCTTGAAGA 1980

QY 1981 TTTGTTTGGAGATTAAGTACTAGCTAGAACACCTTGCGAAGCTTCAATCTTAATGA 2040
 DB 1981 TTTGTTTGGAGATTAAGTACTAGCTAGAACACCTTGCGAAGCTTCAATCTTAATGA 2040

QY 2041 TGAATGGGGCAATGCGCAGTGAAGATGTTAGGCAAGAAAGCCACAGTGAAGATCCAAA 2100
 DB 2041 TGAATGGGGCAATGCGCAGTGAAGATGTTAGGCAAGAAAGCCACAGTGAAGATCCAAA 2100

QY 2101 TAAAGAACTTCAAGAGGAGTGAAGAGCCAGTAGAGAAACAAGTGTGAGCCGAAAGTCCC 2160

Db 2101 TAAGAACTTCAAAAGCGATGAAGAGCCAGTAGAGAAAACCTGCTGAGCCAGAACTCC 2160
Qy 2161 TCAAGTAAAGACTGAAAAAGTGAAGAGCCCACTCAAAAGAGAGAAAGTTTGGCTTCCGA 2220
Db 2161 TCAAGTAAAGACTGAAAAAGTGAAGAGCCCACTCAAAAGAGAGAAAGTTTGGCTTCCGA 2220
Qy 2221 AGTAAAGGATTCAGTCTGAAGAGCCAGTGAAGAGAACTGAGCTGGTTTTCGAAATTA 2280
Db 2221 AGTAAAGGATTCAGTCTGAAGAGCCAGTGAAGAGAACTGAGCTGGTTTTCGAAATTA 2280
Qy 2281 TTTGACTCTTCAAAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 2340
Db 2281 TTTGACTCTTCAAAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 2340
Qy 2341 GTTGTAAAGAGAGATTAATCTTCTCATCTGTAGTAAAGAGAGAGAGAGAGAGAGAGAG 2389
Db 2341 GTTGTAAAGAGAGATTAATCTTCTCATCTGTAGTAAAGAGAGAGAGAGAGAGAGAGAG 2389

RESULT 2

US-09-536-784-55

Sequence 55, Application US/09536784

Patent No. 6573082

GENERAL INFORMATION:

APPLICANT: Choi et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines

NUMBER OF SEQUENCES: 452

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/536,784

FILING DATE: 30-Oct-1997

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/961,083

FILING DATE: OCT-30-1997

ATTORNEY/AGENT INFORMATION:

NAME: Michelle S. Marks

REGISTRATION NUMBER: 41,971

REFERENCE/DOCKET NUMBER: PB340P3

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 55:

SEQUENCE CHARACTERISTICS:

LENGTH: 2389 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 55:

US-09-536-784-55

Query Match 100.0%; Score 2389; DB 4; Length 2389;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTTCTACAGTGGAGCTGTATCAAGCTAGAGAGGTTAAGGAAATTAATCGTGTCTTA 60
Db 1 TTTCTACAGTGGAGCTGTATCAAGCTAGAGAGGTTAAGGAAATTAATCGTGTCTTA 60
Qy 61 TATAGATGAAAAACAAGCAGCAAAAAACGAGAAATTGACTCCTGATGAGGTTAGCA 120
Db 1 TATAGATGAAAAACAAGCAGCAGCAAAAAACGAGAAATTGACTCCTGATGAGGTTAGCA 120

Db 61 TATAGATGAAAAACAAGCAGCAGCAAAAAACGAGAAATTGACTCCTGATGAGGTTAGCA 120
Qy 121 GGTGAAGGAATCAATGCTGAGCAAAATCTGTATCAAGATTAACAGACCAAGGCTATGTAC 180
Db 121 GGTGAAGGAATCAATGCTGAGCAAAATCTGTATCAAGATTAACAGACCAAGGCTATGTAC 180
Qy 181 TTTACATGCGCAGCACTATCATTTATTAATGATGATGATGATGATGATGATGATGATGATG 240
Db 181 TTTACATGCGCAGCACTATCATTTATTAATGATGATGATGATGATGATGATGATGATGATG 240
Qy 241 TGAAGATTTACTATGAAAGATCCAAACTATAGCTAAAGATGAGATATTTGTAATGA 300
Db 241 TGAAGATTTACTATGAAAGATCCAAACTATAGCTAAAGATGAGATATTTGTAATGA 300
Qy 301 GGTCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 360
Db 301 GGTCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 360
Qy 361 TGCCACGCGGATTAAGTCCGTACAAAAGAGAAATGATGACAAAACAAAGAGCATAG 420
Db 361 TGCCACGCGGATTAAGTCCGTACAAAAGAGAAATGATGACAAAACAAAGAGCATAG 420
Qy 421 TCAACATGCTGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 480
Db 421 TCAACATGCTGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 480
Qy 481 AGAGGCTATGCTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540
Db 481 AGAGGCTATGCTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540
Qy 541 TGTGATGCTTATATGCTTCTCTCATGAGATCATTTACATTAATCTTAAGATGAGTT 600
Db 541 TGTGATGCTTATATGCTTCTCTCATGAGATCATTTACATTAATCTTAAGATGAGTT 600
Qy 601 ATCAAGCTAGCGAGTGGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
Db 601 ATCAAGCTAGCGAGTGGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
Qy 661 TTTCAAGAACCTATGCGGACAAATAGCGATTAACATTTCAAGAACAACTGGATACCTTC 720
Db 661 TTTCAAGAACCTATGCGGACAAATAGCGATTAACATTTCAAGAACAACTGGATACCTTC 720
Qy 721 TGTAAAGCATCAGAGACTACAAATACAAACAAAGCAACAAAGCAACAAAGCAACAAAGCA 780
Db 721 TGTAAAGCATCAGAGACTACAAATACAAACAAAGCAACAAAGCAACAAAGCAACAAAGCA 780
Qy 781 AGCAAGTCAAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
Db 781 AGCAAGTCAAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
Qy 841 TCAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
Db 841 TCAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
Qy 901 AGCTAGAGGTTGTCAGTGCACACAGAGATCATTAACATTCCTTACTCTCAAT 960
Db 901 AGCTAGAGGTTGTCAGTGCACACAGAGATCATTAACATTCCTTACTCTCAAT 960
Qy 961 GTCTGAATGGAAGAGAAATCGCTGATATATTCCTTCTGTTATGTTTCAAAACATTTG 1020
Db 961 GTCTGAATGGAAGAGAAATCGCTGATATATTCCTTCTGTTATGTTTCAAAACATTTG 1020
Qy 1021 GGTACAGATTCAGAGGCGAGAACAAAGCCAGCAACGAGCTCGGAGACCTAGTCCAGG 1080
Db 1021 GGTACAGATTCAGAGGCGAGAACAAAGCCAGCAACGAGCTCGGAGACCTAGTCCAGG 1080
Qy 1081 CCCGCAACCTGACCAATATCTTAAATAGACTCAAAATTCCTTCTTGGTATGCTGAGTGT 1140
Db 1081 CCCGCAACCTGACCAATATCTTAAATAGACTCAAAATTCCTTCTTGGTATGCTGAGTGT 1140
Qy 1141 ACGAAAAGTTGGGAGAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1200
Db 1141 ACGAAAAGTTGGGAGAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1200


```

Db      ||||| 540 AGGAGCTACTACTACAGATGATGATTATATCTTTAATCTCTGATATCATATAGAGATAC 599
Qy      ||||| 541 TGTGATGCTTATATATCGTTCCTCATGAGATCATATACATTAATCTCTAAGATATG 600
Db      ||||| 600 TGGTATGCTTATATATCGTTCCTCATGAGATCATATACATTAATCTCTAAGATATG 659
Qy      ||||| 601 ATCAGCTAGCGAGTGTGCTGTCAGAAAGCCTTCTATCTGGTTCGAGGAAATCTGTCAA 660
Db      ||||| 660 ATCAGCTAGCGAGTGTGCTGTCAGAAAGCCTTCTATCTGGTTCGAGGAAATCTGTCAA 719
Qy      ||||| 661 TTCAAGAACTTATCGCCGACAAATAGCGATPACCTTCAAGAACAACTGGGATCCTTC 720
Db      ||||| 720 TTCAAGAACTTATCGCCGACAAATAGCGATPACCTTCAAGAACAACTGGGATCCTTC 779
Qy      ||||| 721 TGTAGCAATCCAGGAATACAAATATCTAACACAAAGCAACACATTAACAGTCA 780
Db      ||||| 780 TGTAGCAATCCAGGAATACAAATATCTAACACAAAGCAACACATTAACAGTCA 839
Qy      ||||| 781 AGCAAGTCAAGATATGACATTTGATAGTCTCTTGAAGACGCTCTCAAACTGCCCTTGA 840
Db      ||||| 840 AGCAAGTCAAGATATGACATTTGATAGTCTCTTGAAGACGCTCTCAAACTGCCCTTGA 899
Qy      ||||| 841 TCAACGACATGATAGATCTGATGCGCTTGTCTTGTATCCAGCAAAATCACAAGTCAAC 900
Db      ||||| 900 TCAACGACATGATAGATCTGATGCGCTTGTCTTGTATCCAGCAAAATCACAAGTCAAC 959
Qy      ||||| 901 AGCTAGAGTGTGAGTGCAGGCCACAGGAGATCATTAACATCTTCCCTTACTCTCAAT 960
Db      ||||| 960 AGCTAGAGTGTGAGTGCAGGCCACAGGAGATCATTAACATCTTCCCTTACTCTCAAT 1019
Qy      ||||| 961 GTCTGAATTTGAGAGAACAAATCGCTCGATATATTCCTTCTGTTATCTGTTCAACATTTG 1020
Db      ||||| 1020 GTCTGAATTTGAGAGAACAAATCGCTCGATATATTCCTTCTGTTATCTGTTCAACATTTG 1079
Qy      ||||| 1021 GGTATCCAGATTCAGAGCCAGAACCAAGTCCACAACTCCGGAACCTTATGTCAGG 1080
Db      ||||| 1080 GGTATCCAGATTCAGAGCCAGAACCAAGTCCACAACTCCGGAACCTTATGTCAGG 1139
Qy      ||||| 1081 CCCCAGACCTGACCAAAATCTTAAATATGATCAAAATCTTCTGTTATCTGTTCAACATTTG 1140
Db      ||||| 1140 CCCCAGACCTGACCAAAATCTTAAATATGATCAAAATCTTCTGTTATCTGTTCAACATTTG 1199
Qy      ||||| 1141 AGCAAAAGTGTGGGAGAGATATGTTTGAAGAAAGGCAATCTCTGTTATGTTCTTGC 1200
Db      ||||| 1200 AGCAAAAGTGTGGGAGAGATATGTTTGAAGAAAGGCAATCTCTGTTATGTTCTTGC 1259
Qy      ||||| 1201 GAAAGATTTACATCTGAAACCTGTTAAATCTTGAAGCAAGTTATCAAAACAGAGAG 1260
Db      ||||| 1260 GAAAGATTTACATCTGAAACCTGTTAAATCTTGAAGCAAGTTATCAAAACAGAGAG 1319
Qy      ||||| 1261 TGTTCACACACTTAACTGCTAAAGAAAGAAATGTTCTCTCTGTCGACAAAGATTTTA 1320
Db      ||||| 1320 TGTTCACACACTTAACTGCTAAAGAAAGAAATGTTCTCTCTGTCGACAAAGATTTTA 1379
Qy      ||||| 1321 TGATTAAGCATATATATCTGTTAACTGAGGCTCATTAAGCCTGTTTGAATTAAGGGTGC 1380
Db      ||||| 1380 TGATTAAGCATATATATCTGTTAACTGAGGCTCATTAAGCCTGTTTGAATTAAGGGTGC 1439
Qy      ||||| 1381 TAAATTCGATTTCCAAAGCCTTGAACAAATTTATGAACGCTTGAATGATGACATTA 1440
Db      ||||| 1440 TAAATTCGATTTCCAAAGCCTTGAACAAATTTATGAACGCTTGAATGATGACATTA 1499
Qy      ||||| 1441 TAAAGAAAAATTTGATGATGATTTATTTGGCATTTCTAGACCAATTAACCAAGAGCG 1500
Db      ||||| 1500 TAAAGAAAAATTTGATGATGATTTATTTGGCATTTCTAGACCAATTAACCAAGAGCG 1559
Qy      ||||| 1501 ACTTGCAAAACCAATTTCTCAAAATGAGTATACGAAGCAAGTGTGATTTGCTCAAT 1560
Db      ||||| 1560 ACTTGCAAAACCAATTTCTCAAAATGAGTATACGAAGCAAGTGTGATTTGCTCAAT 1619
Qy      ||||| 1561 AGCTGATAGATACACAGTCAAGATGTTTACATTTTGTATGAACATGATATATACATGA 1620

```

```

Db      ||||| 1620 AGCTGATAGATACACAGTCAAGATGTTTACATTTTGTATGATATATATACATGA 1679
Qy      ||||| 1621 TGAAGAGATGACATATGTAACGCTCATATGAGGCTATGTCATGATTTGGAAGATAG 1680
Db      ||||| 1680 TGAAGAGATGACATATGTAACGCTCATATGAGGCTATGTCATGATTTGGAAGATAG 1739
Qy      ||||| 1681 CTTTCTGATTAAGAAAAAGTTGAGCTCAAGCCTATATTAAGAAAAAGGATCTTACC 1740
Db      ||||| 1740 CTTTCTGATTAAGAAAAAGTTGAGCTCAAGCCTATATTAAGAAAAAGGATCTTACC 1799
Qy      ||||| 1741 TCCATCTTCAGACCCAGATGTTAAAGCAATCCAACTGAGATATGTCAGACATTTTA 1800
Db      ||||| 1800 TCCATCTTCAGACCCAGATGTTAAAGCAATCCAACTGAGATATGTCAGACATTTTA 1859
Qy      ||||| 1801 CAATCGTGTGAAGGGGAAAAAGAAATTCACCTGCTGACCTTGCATATATATGTTGACA 1860
Db      ||||| 1860 CAATCGTGTGAAGGGGAAAAAGAAATTCACCTGCTGACCTTGCATATATATGTTGACA 1919
Qy      ||||| 1861 TACAGTTAGGTTAAAAACGGTAATTTGATTAATTCCTCATTAAGATCATTAATAT 1920
Db      ||||| 1920 TACAGTTAGGTTAAAAACGGTAATTTGATTAATTCCTCATTAAGATCATTAATAT 1979
Qy      ||||| 1921 TAAATTTGCTTGTGTTGATGATCAACATATACAAAGCTCCAAATGCTATATCTTGAAGA 1980
Db      ||||| 1980 TAAATTTGCTTGTGTTGATGATCAACATATACAAAGCTCCAAATGCTATATCTTGAAGA 2039
Qy      ||||| 1981 TTTGTTGCGACGATTAAGTACTAGTAAACACCTGAGAGAGCTCCACATTTAATGA 2040
Db      ||||| 2040 TTTGTTGCGACGATTAAGTACTAGTAAACACCTGAGAGAGCTCCACATTTAATGA 2099
Qy      ||||| 2041 TGTATGGGGCAATGCGAGTGCATGTTTGAAGCAAAAGACCCAGAGTAAAGATCCAAA 2100
Db      ||||| 2100 TGTATGGGGCAATGCGAGTGCATGTTTGAAGCAAAAGACCCAGAGTAAAGATCCAAA 2159
Qy      ||||| 2101 TAAAGACTTCAAGCGGATGTAAGAGCCAGTBAAGAAAAACCTTCTGAGCCAGAGTCCC 2160
Db      ||||| 2160 TAAAGACTTCAAGCGGATGTAAGAGCCAGTBAAGAAAAACCTTCTGAGCCAGAGTCCC 2219
Qy      ||||| 2161 TCAAGTAGAGCTGAAAAAGTAGAAGCCCAACTCAAGAAAGACAGAGTTTGTCTGCGAA 2220
Db      ||||| 2220 TCAAGTAGAGCTGAAAAAGTAGAAGCCCAACTCAAGAAAGACAGAGTTTGTCTGCGAA 2279
Qy      ||||| 2221 AGTAACGATTTTATGCTGAAAGCCAAATGCAACAGAAATCTTACTGTTTACGAAATTA 2280
Db      ||||| 2280 AGTAACGATTTTATGCTGAAAGCCAAATGCAACAGAAATCTTACTGTTTACGAAATTA 2339
Qy      ||||| 2281 TTTGACTCTTCAATTTATGATTAACAATAGTATATGCGAGACAGAAAAATTAATTGCC 2340
Db      ||||| 2340 TTTGACTCTTCAATTTATGATTAACAATAGTATATGCGAGACAGAAAAATTAATTGCC 2399
Qy      ||||| 2341 GTTGTAAAGGAAGTATCTTCACTGTATAGTAAAGAAAAATTAAC 2389
Db      ||||| 2400 GTTGTAAAGGAAGTATCTTCACTGTATAGTAAAGAAAAATTAAC 2448

```

```

RESULT 4
US-08-961-527-94
; Sequence 94, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33

```

OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ. ID NO: 94:
SEQUENCE CHARACTERISTICS:
LENGTH: 8195 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-94

Query Match 97.9%; Score 2338; DB 4; Length 8195;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTCTTACGAGTTGGAGCTGTATTAACCTGGAAGGTTAAGAAATTAATCGTCTTCTTA 60
DB 3053 TTCTTACGAGTTGGAGCTGTATTAACCTGGAAGGTTAAGAAATTAATCGTCTTCTTA 3112
QY 61 TTTAGATGAAAAACAAGCGACGCAAAAAAGGAAATTTGACCTCGATGAGGTTAGCA 120
DB 3113 TTTAGATGAAAAACAAGCGACGCAAAAAAGGAAATTTGACCTCGATGAGGTTAGCA 3172
QY 121 GCGTGAAGAAATCAATGCTGAGCAAAATGCTCATCAAGATPAACAGCAAGGCTATGTCAC 180
DB 3173 GCGTGAAGAAATCAATGCTGAGCAAAATGCTCATCAAGATPAACAGCAAGGCTATGTCAC 3232
QY 181 TTTCACATGGCGACCACTATCATTTATTAACAATGGTAAAGTTCTTATGAGGCTATCATCAG 240
DB 3233 TTTCACATGGCGACCACTATCATTTATTAACAATGGTAAAGTTCTTATGAGGCTATCATCAG 3292
QY 241 TGAAGAAATTAACATGAAGAAATCAAACTATTAAGTAAAGATGAGGATATTGTTAATA 300
DB 3293 TGAAGAAATTAACATGAAGAAATCAAACTATTAAGTAAAGATGAGGATATTGTTAATA 3352
QY 301 GGTCAAGGCTGATATGTTATCAAGGTAGATGGAATACTATGTTTACCTTAAGGATGC 360
DB 3353 GGTCAAGGCTGATATGTTATCAAGGTAGATGGAATACTATGTTTACCTTAAGGATGC 3412
QY 361 TGGCCACGGGATTAAGTCCGTTCAAAAAGGAATCAATCGCAAAAAACAAGAGATAG 420
DB 3413 TGGCCACGGGATTAAGTCCGTTCAAAAAGGAATCAATCGCAAAAAACAAGAGATAG 3472
QY 421 TCAACATCGTGAAGGGAACCTCAAGAAACGATGAGTGGCTTGGCCATGCTTCGCA 480
DB 3473 TCAACATCGTGAAGGGAACCTCAAGAAACGATGAGTGGCTTGGCCATGCTTCGCA 3532
QY 481 AGGACGCTATACTACAGATGATGTTATATCTTTAATGCTTCTGATATCAATAGAGATAC 540
DB 3533 AGGACGCTATACTACAGATGATGTTATATCTTTAATGCTTCTGATATCAATAGAGATAC 3592
QY 541 TGGTATGCTTATATGTTCTCTCATGAGATCAATTAACATTAATTCCTTAAGATAGTT 600
DB 3593 TGGTATGCTTATATGTTCTCTCATGAGATCAATTAACATTAATTCCTTAAGATAGTT 3652
QY 601 ATCAGTACGAGTTGGCTGTCAGAAAGCTTCCATCTGGTCGAGAAATTCGTCAA 660
DB 3653 ATCAGTACGAGTTGGCTGTCAGAAAGCTTCCATCTGGTCGAGAAATTCGTCAA 3712
QY 661 TTCAGAACTTATCGCCGACAAATATAGCATTAACATTCAGAACTGCGGTACTTTC 720

DB 3713 TTCAAGAACTTATCGCCGACAAATATAGCATTAACACTTCAGAACTGGGTACTTTC 3772
QY 721 TGTAAAGCAATCCAGAACTACAAATATCAACCAAGCAACACACACATCAACAGTCA 780
DB 3773 TGTAAAGCAATCCAGAACTACAAATATCAACCAAGCAACACACACATCAACAGTCA 3832
QY 781 AGCAAGTCAAGTAAATGACATTAATGATGCTCTTGAACAGCTCTCAAACTGCTTTGAG 840
DB 3833 AGCAAGTCAAGTAAATGACATTAATGATGCTCTTGAACAGCTCTCAAACTGCTTTGAG 3892
QY 841 TCAAGCAATGTAGATCTGATGCTTGTCTTGAATCCAGCAAAATCAACAGTCAAC 900
DB 3893 TCAAGCAATGTAGATCTGATGCTTGTCTTGAATCCAGCAAAATCAACAGTCAAC 3952
QY 901 AGCTAGAGTGTGAGGAGCCACAGGAGATCATTAACACTTCAATCCCTACTCTCAAT 960
DB 3953 AGCTAGAGTGTGAGGAGCCACAGGAGATCATTAACACTTCAATCCCTACTCTCAAT 4012
QY 961 GTCTGAATTTGAAGAAGCAATGCTCTGTAATTAATCCCTCTGTTATCGTTCAACCATG 1020
DB 4013 GTCTGAATTTGAAGAAGCAATGCTCTGTAATTAATCCCTCTGTTATCGTTCAACCATG 4072
QY 1021 GGTACCAATTCAGGCGCAAGCAACCAAGTCCACACCGACTCCGGAACCTAGTCCAG 1080
DB 4073 GGTACCAATTCAGGCGCAAGCAACCAAGTCCACACCGACTCCGGAACCTAGTCCAG 4132
QY 1081 CCGGCACTGACCAAAATCTTAAATATGACTCAAAATCTTCTTGGTATGACAGCTGGT 1140
DB 4133 CCGGCACTGACCAAAATCTTAAATATGACTCAAAATCTTCTTGGTATGACAGCTGGT 4192
QY 1141 AGCAAAAGTTGGGGAAGGATATGTAATGCAAGAAAGGCACTCTCGTTATGTCCTTGC 1200
DB 4193 AGCAAAAGTTGGGGAAGGATATGTAATGCAAGAAAGGCACTCTCGTTATGTCCTTGC 4252
QY 1201 GAAAGATTTACCATGTGAAACTGTTAAATCTTGAAGCAAGTATCAAAACAAGAG 1260
DB 4253 GAAAGATTTACCATGTGAAACTGTTAAATCTTGAAGCAAGTATCAAAACAAGAG 4312
QY 1261 TGTTCACACACTTTAATCTGCTTAAAGAAAGAAATGTTGCTCTCGTGAACCAAGATTTA 1320
DB 4313 TGTTCACACACTTTAATCTGCTTAAAGAAAGAAATGTTGCTCTCGTGAACCAAGATTTA 4372
QY 1321 TGATTAAGCAATTAATCTGTTAATGAGGCTCATTAAGGCTTGTGNAATTAAGGCTG 1380
DB 4373 TGATTAAGCAATTAATCTGTTAATGAGGCTCATTAAGGCTTGTGNAATTAAGGCTG 4432
QY 1381 TAATCTGATTTCAAGCCTTAGACAAATTAATTAAGCGCTTGAATGATGAATCGACTAA 1440
DB 4433 TAATCTGATTTCAAGCCTTAGACAAATTAATTAAGCGCTTGAATGATGAATCGACTAA 4492
QY 1441 TAAAGAAAAATGATGATGATTTATGTCATTTCTAGACCAATTAATCCATCCAGAGG 1500
DB 4493 TAAAGAAAAATGATGATGATTTATGTCATTTCTAGACCAATTAATCCATCCAGAGG 4552
QY 1501 ACTTGCAAAACCAATTTCTCAATTAAGATGATCTGAAGCAAGTTCGATTTGCTCAAT 1560
DB 4553 ACTTGCAAAACCAATTTCTCAATTAAGATGATCTGAAGCAAGTTCGATTTGCTCAAT 4612
QY 1561 AGCTGTAAGTATTAACAGCTGATGATTTATTTGATGAACATGATATTAATCACTGA 1620
DB 4613 AGCTGTAAGTATTAACAGCTGATGATTTATTTGATGAACATGATATTAATCACTGA 4672
QY 1621 TGAAGAGATGATATGTAAGCTCTCATATGAGGCAATGATCACTGATTTGGAAGATAG 1680
DB 4673 TGAAGAGATGATATGTAAGCTCTCATATGAGGCAATGATCACTGATTTGGAAGATAG 4732
QY 1681 CCTTCTGATTAAGAAAAAGTTCAGCTCAACCTTATTAAGAAAAAGATCTTAC 1740
DB 4733 CCTTCTGATTAAGAAAAAGTTCAGCTCAACCTTATTAAGAAAAAGATCTTAC 4792
QY 1741 TCCATCTCCAGACGAGATGTTAAAGCAATCAACTGAGATAGTGCAGCACTATTTTA 1800

Db 4793 TCACATCTCCAGACGACAGATGTTAAAGCAATCCAACTGGAGATAGTGACAGACTATTTA 4852
Qy 1801 CAATCGTGTGAAGGGGAAAAAGCAATTCACCTGCTGACTTCATATATATGTTAGACA 1860
Db 4853 CAATGCTGTGAAGGGGAAAAAGCAATTCACCTGCTGACTTCATATATGTTAGACA 4912
Qy 1861 TACAGTTAGAGTTAAAAAGGTAATTTGATATATTCCTGATAGATGATCAATCAATATAT 1920
Db 4913 TACAGTTAGAGTTAAAAAGGTAATTTGATATATTCCTGATAGATGATCAATCAATATAT 4972
Qy 1921 TAAATTTGCTGTTGTTGATGATCAACATACAAAGCTCCAAATGCTATACCTTGAAGA 1980
Db 4973 TAAATTTGCTGTTGTTGATGATCAACATACAAAGCTCCAAATGCTATACCTTGAAGA 5032
Qy 1981 TTTGTTTCGACGAGTTAAGTACTAGTGAACACCCCTGACGAACTCCACATTTCAATGA 2040
Db 5032 TTTGTTTCGACGAGTTAAGTACTAGTGAACACCCCTGACGAACTCCACATTTCAATGA 5092
Qy 2041 TGGATGGGGCAATGCGACATGATGTTAGGCAAGAAAGCAACAGTGAAGATCCAAA 2100
Db 5093 TGGATGGGGCAATGCGACATGATGTTAGGCAAGAAAGCAACAGTGAAGATCCAAA 5152
Qy 2101 TAAAGACTTCAAAGCGGATGAAGACCGATGAGAGAAACCTGCTGACCGCAAGTCCC 2160
Db 5153 TAAAGACTTCAAAGCGGATGAAGACCGATGAGAGAAACCTGCTGACCGCAAGTCCC 5212
Qy 2161 TCAAGTAGAGACTGAAAAAGTGAAGCCCACTCAAGAGAGAGAGTTTGGCTTGGCAA 2220
Db 5213 TCAAGTAGAGACTGAAAAAGTGAAGCCCACTCAAGAGAGAGAGTTTGGCTTGGCAA 5272
Qy 2221 AGTAAAGGATTTCTAGCTGAAAAGCCCAATGCAACAGAACTAGCTGTTTACGAATTA 2280
Db 5273 AGTAAAGGATTTCTAGCTGAAAAGCCCAATGCAACAGAACTAGCTGTTTACGAATTA 5332
Qy 2281 TTTGACTCTTCAAAATTAATGAGATACAAATATGATCTGCGACAGAGAGAAATTTCTTGC 2340
Db 5333 TTTGACTCTTCAAAATTAATGAGATACAAATATGATCTGCGACAGAGAGAAATTTCTTGC 5392
Qy 2341 GTTGTAAAGAGTAATCTCTCATCTGTAGTAGGAAAAAATAAAC 2389
Db 5393 GTTGTAAAGAGTAATCTCTCATCTGTAGTAGGAAAAAATAAAC 5441

RESULT 5
US-08-961-083-65
; Sequence 65, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961.083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373

; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2290 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-083-65

Query Match 2.2%; Score 53; DB 3; Length 2290;
Best Local Similarity 100.0%; Pred. No. 1.3e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 983 GCTGTATTATTCCTTCGTTATCGTTCAACCATTTGGTACGAGATTCAAG 1035
Db 944 GCTGTATTATTCCTTCGTTATCGTTCAACCATTTGGTACGAGATTCAAG 996

RESULT 6
US-09-536-784-65
; Sequence 65, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536.784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961.083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2290 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-09-536-784-65

Query Match 2.2%; Score 53; DB 4; Length 2290;
Best Local Similarity 100.0%; Pred. No. 1.3e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 983 GCTGTATTATTCCTTCGTTATCGTTCAACCATTTGGTACGAGATTCAAG 1035
Db 944 GCTGTATTATTCCTTCGTTATCGTTCAACCATTTGGTACGAGATTCAAG 996

RESULT 7
US-08-961-527-243
; Sequence 243, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunesh
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; INFORMATION FOR SEQ ID NO: 243:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2359 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-527-243
Query Match 2.2%; Score 53; DB 4; Length 2359;
Best Local Similarity 100.0%; Pred. No. 1.3e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 983 GCTGATATTATTCCTTCGTTATGTTCAACCACTGGGTACCAAGATTCAAG 1035
DB 1879 GCTCGATATTATTCCTTCGTTATGTTCAACCACTGGGTACCAAGATTCAAG 1931
RESULT 8
US-09-468-656A-5
; Sequence 5, Application US/09468656A
; Patent No. 6582706
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
; FILE REFERENCE: 469201-444
; CURRENT APPLICATION NUMBER: US/09/468,656A
; CURRENT FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 2531
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; US-09-468-656A-5

Query Match 2.2%; Score 53; DB 4; Length 2531;
Best Local Similarity 100.0%; Pred. No. 1.3e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 983 GCTGATATTATTCCTTCGTTATGTTCAACCACTGGGTACCAAGATTCAAG 1035
DB 1003 GCTCGATATTATTCCTTCGTTATGTTCAACCACTGGGTACCAAGATTCAAG 1055
RESULT 9
US-09-468-656A-11
; Sequence 11, Application US/09468656A
; Patent No. 6582706
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
; FILE REFERENCE: 469201-444
; CURRENT APPLICATION NUMBER: US/09/468,656A
; CURRENT FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 2531
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; US-09-468-656A-11
Query Match 2.2%; Score 53; DB 4; Length 2531;
Best Local Similarity 100.0%; Pred. No. 1.3e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 983 GCTGATATTATTCCTTCGTTATGTTCAACCACTGGGTACCAAGATTCAAG 1035
DB 1000 GCTCGATATTATTCCTTCGTTATGTTCAACCACTGGGTACCAAGATTCAAG 1052
RESULT 10
US-08-961-527-355
; Sequence 355, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunesh
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1

TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 355:
SEQUENCE CHARACTERISTICS:
LENGTH: 973 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-355

Query Match 1.8%; Score 44; DB 4; Length 973;
Best Local Similarity 100.0%; Pred. No. 4,1e-12;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 572 CATTACCATTCATCCAGAGATGATTCAGTACGAGCTT 615
Db 722 CATTACCATTCATTCCTAAGAGATGATTCAGTACGAGCTT 765

RESULT 11

US-08-961-083-282/c
Sequence 282; Application US/08961083
Patent No. 6159469

GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:

ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2

TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 282:
SEQUENCE CHARACTERISTICS:
LENGTH: 40 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-083-282

Query Match 1.3%; Score 30; DB 3; Length 40;
Best Local Similarity 100.0%; Pred. No. 3.8e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2360 CCTTCATCTGTAGTAAAGAAAAATAAAC 2389
Db 40 CCTTCATCTGTAGTAAAGAAAAATAAAC 11

RESULT 12
US-09-536-784-282/c

Sequence 282; Application US/09536784
Patent No. 6573082

GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/536,784
FILING DATE: 30-Oct-1997
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: OCT-30-1997

ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB340P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 282:
SEQUENCE CHARACTERISTICS:
LENGTH: 40 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 282:
US-09-536-784-282

Query Match 1.3%; Score 30; DB 4; Length 40;
Best Local Similarity 100.0%; Pred. No. 3.8e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2360 CCTTCATCTGTAGTAAAGAAAAATAAAC 2389
Db 40 CCTTCATCTGTAGTAAAGAAAAATAAAC 11

RESULT 13
US-09-468-656A-3/c
Sequence 3; Application US/09468656A
Patent No. 6582706

GENERAL INFORMATION:
APPLICANT: Johnson, Leslie S.
TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
FILE REFERENCE: 469201-444
CURRENT APPLICATION NUMBER: US/09/468,656A
CURRENT FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: 60/113,048
PRIOR FILING DATE: 1998-12-21
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO: 3
LENGTH: 40

TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Reverse primer

OTHER INFORMATION: used in amplification of the Sp36 gene sequence.
US-09-468-656A-3

Query Match
Best Local Similarity 100.0%; Score 30; DB 4; Length 40;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2360 CCTTCATCTGTAGTAAGGAAAAATTAAC 2389
DB 40 CCTTCATCTGTAGTAAGGAAAAATTAAC 11

RESULT 14
US-09-468-656A-1
Sequence 1, Application US/09468656A
Patent No. 6582706
GENERAL INFORMATION:
APPLICANT: Johnson, Leslie S.
APPLICANT: Adamou, John E.
TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
FILE REFERENCE: 469201-444
CURRENT APPLICATION NUMBER: US/09/468,656A
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: 60/113,048
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 36
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Forward primer
OTHER INFORMATION: used in amplification of the Sp36 gene sequence.
US-09-468-656A-1

Query Match
Best Local Similarity 100.0%; Score 27; DB 4; Length 36;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTTACGAGTTGGAGCTGTATCAAGC 27
DB 10 TTCTTACGAGTTGGAGCTGTATCAAGC 36

RESULT 15
US-08-961-083-281
Sequence 281, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:

FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 281:
SEQUENCE CHARACTERISTICS:
LENGTH: 37 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-083-281

Query Match
Best Local Similarity 100.0%; Score 27; DB 3; Length 37;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTTACGAGTTGGAGCTGTATCAAGC 27
DB 11 TTCTTACGAGTTGGAGCTGTATCAAGC 37

RESULT 16
US-09-536-784-281
Sequence 281, Application US/09536784
Patent No. 6573062
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/536,784
FILING DATE: 30-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: OCT-30-1997
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB340P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 281:
SEQUENCE CHARACTERISTICS:
LENGTH: 37 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 281:
US-09-536-784-281

Query Match
Best Local Similarity 100.0%; Score 27; DB 4; Length 37;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTTACGAGTTGGAGCTGTATCAAGC 27

Db 11 TTCTTACGAGTTGGACGTATCATCAGC 37

RESULT 17
US-09-468-656A-2
; Sequence 2, Application US/09468656A

; Patent No. 6582706

; GENERAL INFORMATION:

; APPLICANT: Johnson, Leslie S.

; APPLICANT: Adamou, John E.

; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus

; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural

; FILE REFERENCE: 469201-444

; CURRENT APPLICATION NUMBER: US/09/468,656A

; CURRENT FILING DATE: 1999-12-02

; PRIOR APPLICATION NUMBER: 60/113,048

; PRIOR FILING DATE: 1998-12-21

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: Patent In Ver. 2.1

; SEQ ID NO: 2

; LENGTH: 35

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE: OTHER INFORMATION: Description of Artificial Sequence: Forward primer

; OTHER INFORMATION: used in amplification of the sp36 gene sequence.

US-09-468-656A-2

Query Match 1.0%; Score 26; DB 4; Length 35;

Best Local Similarity 100.0%; Pred. No. 0.0038;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 ACTGATCAAGCTAGACGTTAAGC 41

Db 10 ACTGATCAAGCTAGACGTTAAGC 35

RESULT 18
US-08-961-083-181
; Sequence 181, Application US/08961083

; Patent No. 6159469

; GENERAL INFORMATION:

; APPLICANT: Choi et. al.

; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines

; NUMBER OF SEQUENCES: 452

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/961,083

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Brookes, A. Anders

; REGISTRATION NUMBER: 36,373

; REFERENCE/DOCKET NUMBER: PB340P2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (301) 309-8504

; TELEFAX: (301) 309-8512

; INFORMATION FOR SEQ ID NO: 181:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1342 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

US-08-961-083-181

Query Match 1.0%; Score 24; DB 3; Length 1342;

Best Local Similarity 100.0%; Pred. No. 0.042;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 546 ATGCTTATATGTTCTTCATGAG 569

Db 525 ATGCTTATATGTTCTTCATGAG 548

RESULT 19
US-09-536-784-181
; Sequence 181, Application US/09536784

; Patent No. 6573082

; GENERAL INFORMATION:

; APPLICANT: Choi et. al.

; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines

; NUMBER OF SEQUENCES: 452

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/536,784

; FILING DATE: 30-Oct-1997

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/961,083

; FILING DATE: OCT-30-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Michelle S. Marks

; REGISTRATION NUMBER: 41,971

; REFERENCE/DOCKET NUMBER: PB340P3

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (301) 309-8504

; TELEFAX: (301) 309-8512

; INFORMATION FOR SEQ ID NO: 181:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1342 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; SEQUENCE DESCRIPTION: SEQ ID NO: 181:

US-09-536-784-181

Query Match 1.0%; Score 24; DB 4; Length 1342;

Best Local Similarity 100.0%; Pred. No. 0.042;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 546 ATGCTTATATGTTCTTCATGAG 569

Db 525 ATGCTTATATGTTCTTCATGAG 548

RESULT 20

US-09-468-656A-7
; Sequence 7, Application US/09468656A
; Patent No. 6582706

GENERAL INFORMATION:
APPLICANT: Johnson, Leslie S.
APPLICANT: Adamou, John E.
TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
TITLE OF INVENTION: Motifs
FILE REFERENCE: 469201-444
CURRENT APPLICATION NUMBER: US/09/468,656A
CURRENT FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: 60/113,048
PRIOR FILING DATE: 1998-12-21
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 7
LENGTH: 1455
TYPE: DNA
ORGANISM: Streptococcus pneumoniae
US-09-468-656A-7

Query Match
Best Local Similarity 100.0%; Score 24; DB 4; Length 1455;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 546 ATGCTTATATCGTTCTCATGAG 569
DB 596 ATGCTTATATCGTTCTCATGAG 619

RESULT 21
US-08-961-527-192/c
Sequence 192, Application US/08961527
Patent No. 6420135
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION/DOCKET NUMBER: 36,373
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 192:
SEQUENCE CHARACTERISTICS:
LENGTH: 6867 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-192

Query Match
Best Local Similarity 100.0%; Score 24; DB 4; Length 6867;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 546 ATGCTTATATCGTTCTCATGAG 569
DB 6190 ATGCTTATATCGTTCTCATGAG 6167

RESULT 22
US-08-998-416-895
Sequence 895, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtle, Philipp
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPHII
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION/DOCKET NUMBER: 38, 241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8589
INFORMATION FOR SEQ ID NO: 895:
SEQUENCE CHARACTERISTICS:
LENGTH: 622 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1559RP
US-08-998-416-895

Query Match
Best Local Similarity 100.0%; Score 19; DB 3; Length 622;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 698 TCAAGACAACTGGGTAC 716
DB 383 TCAAGACAACTGGGTAC 401

RESULT 23
US-08-212-133A-7
Sequence 7, Application US/08212133A
Patent No. 5663060
GENERAL INFORMATION:
APPLICANT: Lollar, John S.

APPLICANT: Runge, Marschall S.
TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kilpatrick & Cody
STREET: 100 Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: US
ZIP: 30303

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/212,133A
FILING DATE: March 11, 1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,004
FILING DATE: 07-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: EMU/76677
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-572-6508
TELEFAX: 404-572-6555
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 7493 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear.
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 1..407
OTHER INFORMATION: /rpt_type="terminal"
OTHER INFORMATION: /note="5'UTR"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7471..7476
OTHER INFORMATION: /function="PolyA_signal"
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 7368..7493
OTHER INFORMATION: /rpt_type="terminal"
OTHER INFORMATION: /note="3'UTR"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 408..7367
OTHER INFORMATION: /product="Coagulation Factor VIII"
PUBLICATION INFORMATION:
AUTHORS: Elder, F.
AUTHORS: Lakich, D.
AUTHORS: Gitschier, J.
TITLE: Sequence of the Murine Factor VIII cDNA.
Patent No. 5663060
JOURNAL: Genomics
VOLUME: 16
PAGES: 374-379
DATE: 1993
RELEVANT RESIDUES IN SEQ ID NO: 7: FROM 1 TO 7476
US-08-212-133A-7

Query Match 0.8%; Score 19; DB 1; Length 7493;
Best Local Similarity 100.0%; Pred. No. 14;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2347 AAAGAAGTATCTTCA 2365
DB 206 AAAGAAGTATCTTCA 224

RESULT 24
US-08-474-503-5
Sequence 5, Application US/08474503
Patent No. 574446
GENERAL INFORMATION:
APPLICANT: Emory University
TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kilpatrick & Cody
STREET: 1100 Peachtree Street, Suite 2800
CITY: Atlanta
STATE: Georgia
COUNTRY: US
ZIP: 30309

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,503
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pratt, John S.
REGISTRATION NUMBER: 29,476
REFERENCE/DOCKET NUMBER: EMU106CIP(3)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-815-6500
TELEFAX: 404-815-6555
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 7493 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 1..407
OTHER INFORMATION: /rpt_type="terminal"
OTHER INFORMATION: /note="5'UTR"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7471..7476
OTHER INFORMATION: /function="PolyA_signal"
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 7368..7493
OTHER INFORMATION: /rpt_type="terminal"
OTHER INFORMATION: /note="3'UTR"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 408..7367
OTHER INFORMATION: /product="Coagulation Factor VIII"
PUBLICATION INFORMATION:
AUTHORS: Elder, F.
AUTHORS: Lakich, D.
AUTHORS: Gitschier, J.
TITLE: Sequence of the Murine Factor VIII cDNA.
Patent No. 574446

JOURNAL: Genomics
VOLUME: 16
PAGES: 374-379
DATE: 1993
RELEVANT RESIDUES IN SEQ ID NO: 5: FROM 1 TO 7476
US-08-474-503-5

Query Match
Best Local Similarity 100.0%; Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0.8%; Score 19; DB 1; Length 7493;

Qy 2347 AAAAGAGTAATCCTTCA 2365
Db 206 AAAAGAGTAATCCTTCA 224

RESULT 25
US-08-670-707A-5
Sequence 5, Application US/08670707A
Patent No. 5859204
GENERAL INFORMATION:
APPLICANT: Iolliar, John S.
TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
STREET: 5370 Manhattan Circle Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: USA
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/670,707A
FILING DATE: 26-JUN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US94/13200
FILING DATE: 15-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/212,133
FILING DATE: 11-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,004
FILING DATE: 07-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Greenlee, Lorraine L.
REGISTRATION NUMBER: 27,894
REFERENCE/DOCKET NUMBER: 75-95F
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/499-8080
TELEFAX: 303/499-8089
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 7493 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 1..407
OTHER INFORMATION: /rpt_type="terminal"
OTHER INFORMATION: /note="5' UTR"
FEATURE:

NAME/KEY: misc feature
LOCATION: 7471..7476
OTHER INFORMATION: /function="polyA signal"
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 7368..7493
OTHER INFORMATION: /rpt_type="terminal"
OTHER INFORMATION: /note="3' UTR"
FEATURE:
NAME/KEY: misc feature
LOCATION: 408..7367
OTHER INFORMATION: /product="coagulation factor VIII"
PUBLICATION INFORMATION:
AUTHORS: Elder, F.
AUTHORS: Iakich, D.
AUTHORS: Gitschier, J.
TITLE: Sequence of the murine Factor VIII cDNA
Patent No. 5859204
JOURNAL: Genomics
VOLUME: 16
PAGES: 374-379
DATE: 1993
US-08-670-707A-5

Query Match
Best Local Similarity 100.0%; Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0.8%; Score 19; DB 2; Length 7493;

Qy 2347 AAAAGAGTAATCCTTCA 2365
Db 206 AAAAGAGTAATCCTTCA 224

RESULT 26
US-09-037-601-5
Sequence 5, Application US/09037601
Patent No. 6180371
GENERAL INFORMATION:
APPLICANT: Iolliar, John S.
TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
STREET: 5370 Manhattan Circle Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: USA
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/037,601
FILING DATE: 26-JUN-1996
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US94/13200
FILING DATE: 15-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/212,133
FILING DATE: 11-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,004
FILING DATE: 07-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Ferber, Donna M.
REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 75-95F
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/499-8080
TELEFAX: 303/499-8089

```

; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 7493 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: not relevant
;   MOLECULE TYPE: cDNA to mRNA
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   ORIGINAL SOURCE:
;     ORGANISM: Mus musculus
;   FEATURE:
;     NAME/KEY: repeat_unit
;     LOCATION: 1..407
;     OTHER INFORMATION: /rpt_type="terminal"
;     OTHER INFORMATION: /note="5' UTR"
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: 7471..7476
;     OTHER INFORMATION: /function="polyA signal"
;   FEATURE:
;     NAME/KEY: repeat_unit
;     LOCATION: 7368..7493
;     OTHER INFORMATION: /rpt_type="terminal"
;     OTHER INFORMATION: /note="3' UTR"
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: 408..7367
;     OTHER INFORMATION: /product="coagulation factor VIII"
;   PUBLICATION INFORMATION:
;     AUTHORS: Elder, F.
;     AUTHORS: Lakich, D.
;     AUTHORS: Gitschier, J.
;     TITLE: Sequence of the murine Factor VIII cDNA
;     Patent No. 6180371
;     JOURNAL: Genomics
;     VOLUME: 16
;     PAGES: 374-379
;     DATE: 1993
;   US-09-037-601-5

Query Match      0.8%; Score 19; DB 3; Length 7493;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2347 AAAAGAGTAATCCTTCA 2365
Db      206 AAAAGAGTAATCCTTCA 224

RESULT 27
US-09-315-179-5
; Sequence 5, Application US/09315179
; Patent No. 6376463
; GENERAL INFORMATION:
;   APPLICANT: Lollar, John S
;   TITLE OF INVENTION: Modified Factor VIII
;   FILE REFERENCE: 75-95H
;   CURRENT APPLICATION NUMBER: US/09/315,179
;   EARLIER FILING DATE: 1999-05-20
;   EARLIER APPLICATION NUMBER: U.S. 09/037,601
;   EARLIER FILING DATE: 1998-03-10
;   EARLIER APPLICATION NUMBER: U.S. 08/670,707
;   EARLIER FILING DATE: 1996-06-26
;   EARLIER APPLICATION NUMBER: PCT/US97/11155
;   EARLIER FILING DATE: 1997-06-26
;   EARLIER APPLICATION NUMBER: PCT/US94/13200
;   EARLIER FILING DATE: 1994-11-15
;   EARLIER APPLICATION NUMBER: U.S. 08/212,133
;   EARLIER FILING DATE: 1994-03-11
;   EARLIER APPLICATION NUMBER: U.S. 07/864,004
;   EARLIER FILING DATE: 1992-04-07
;   NUMBER OF SEQ ID NOS: 40
```

```

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 5
; LENGTH: 7493
; TYPE: DNA
; ORGANISM: Mus musculus
; US-09-315-179-5

Query Match      0.8%; Score 19; DB 4; Length 7493;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2347 AAAAGAGTAATCCTTCA 2365
Db      206 AAAAGAGTAATCCTTCA 224

RESULT 28
PCT-US94-13200-5
; Sequence 5, Application PC/TUS9413200
; GENERAL INFORMATION:
;   APPLICANT: Emory University
;   TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
;   NUMBER OF SEQUENCES: 12
;   CORRESPONDENCE ADDRESS:
;     ADDRESS: Kilpatrick & Cody
;     STREET: 1100 Peachtree Street, Suite 2800
;     CITY: Atlanta
;     STATE: Georgia
;     COUNTRY: US
;     ZIP: 30309
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: PCT/US94/13200
;     FILING DATE: 15-NOV-1994
;     CLASSIFICATION:
;       ATTORNEY/AGENT INFORMATION:
;         NAME: Pabst, Patricia L.
;         REGISTRATION NUMBER: 31,284
;         REFERENCE/DOCKET NUMBER: EMU106CIP(2)
;         TELECOMMUNICATION INFORMATION:
;           TELEPHONE: 404-815-6508
;           TELEFAX: 404-815-6555
;   INFORMATION FOR SEQ ID NO: 5:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 7493 base pairs
;       TYPE: nucleic acid
;       STRANDEDNESS: single
;       TOPOLOGY: linear
;       MOLECULE TYPE: cDNA to mRNA
;       HYPOTHETICAL: NO
;       ANTI-SENSE: NO
;       ORIGINAL SOURCE:
;         ORGANISM: Mus musculus
;       FEATURE:
;         NAME/KEY: repeat_unit
;         LOCATION: 1..407
;         OTHER INFORMATION: /rpt_type="terminal"
;         OTHER INFORMATION: /note="5' UTR"
;       FEATURE:
;         NAME/KEY: misc_feature
;         LOCATION: 7471..7476
;         OTHER INFORMATION: /function="polyA signal"
;       FEATURE:
;         NAME/KEY: repeat_unit
;         LOCATION: 7368..7493
;         OTHER INFORMATION: /rpt_type="terminal"
;         OTHER INFORMATION: /note="3' UTR"
;       FEATURE:
;         NAME/KEY: misc_feature
```

LOCATION: 408..7367
OTHER INFORMATION: /product= "Coagulation Factor VIII"
PUBLICATION INFORMATION:
AUTHORS: Elder, F.
AUTHORS: Lakich, D.
AUTHORS: Gitschier, J.
TITLE: Sequence of the Murine Factor VIII cDNA.
JOURNAL: Genomics
VOLUME: 16
PAGES: 374-379
DATE: 1993
RELEVANT RESIDUES IN SEQ ID NO: 5: FROM 1 TO 7476
PCT-US94-13200-5

Query Match 0.8%; Score 19; DB 5; Length 7493;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2347 AAAGAGTAATCCTTCA 2365
DB 206 AAAGAGTAATCCTTCA 224

RESULT 29
US-09-641-638-52/c
Sequence 52, Application US/09641638
Patent No. 6432648
GENERAL INFORMATION:
APPLICANT: Blumenfeld, Marta
APPLICANT: Bougueleret, Lydie
APPLICANT: Chumakov, Ilya
APPLICANT: Cohen, Annick
TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
FILE REFERENCE: GENSET.051CPI
CURRENT APPLICATION NUMBER: US/09/641,638
PRIOR FILING DATE: 2000-08-16
PRIOR APPLICATION NUMBER: US 09/502,330
PRIOR FILING DATE: 2000-02-11
PRIOR APPLICATION NUMBER: US 60/133,200
PRIOR FILING DATE: 1999-05-07
PRIOR APPLICATION NUMBER: US 09/275,267
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: US 60/119,917
PRIOR FILING DATE: 1999-02-12
NUMBER OF SEQ ID NOS: 1304
SOFTWARE: Patent.pm
SEQ ID NO 52
LENGTH: 499
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: allele
LOCATION: 466
OTHER INFORMATION: 12-215-467 : polymorphic base G or T
NAME/KEY: misc_binding
LOCATION: 446_465
OTHER INFORMATION: 12-215-467.misl, potential
NAME/KEY: misc_binding
LOCATION: 467..486
OTHER INFORMATION: 12-215-467.mis2, potential complement
NAME/KEY: primer_bind
LOCATION: 1..20
OTHER INFORMATION: upstream amplification primer
NAME/KEY: primer_bind
LOCATION: 479..499
OTHER INFORMATION: downstream amplification primer, complement
NAME/KEY: misc_binding
LOCATION: 454_478
OTHER INFORMATION: 12-215-467 potential probe
NAME/KEY: misc_feature
LOCATION: 87
OTHER INFORMATION: n=a, g, c or t

US-09-641-638-52

Query Match 0.8%; Score 18; DB 4; Length 499;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 751 CACAGCAACACAGCAA 768
DB 76 CACAGCAACACAGCAA 59

RESULT 30
US-09-641-638-572/c
Sequence 572, Application US/09641638
Patent No. 6432648
GENERAL INFORMATION:
APPLICANT: Blumenfeld, Marta
APPLICANT: Bougueleret, Lydie
APPLICANT: Chumakov, Ilya
APPLICANT: Cohen, Annick
TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
FILE REFERENCE: GENSET.051CPI
CURRENT APPLICATION NUMBER: US/09/641,638
PRIOR FILING DATE: 2000-08-16
PRIOR APPLICATION NUMBER: US 09/502,330
PRIOR FILING DATE: 2000-02-11
PRIOR APPLICATION NUMBER: US 60/133,200
PRIOR FILING DATE: 1999-05-07
PRIOR APPLICATION NUMBER: US 09/275,267
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: US 60/119,917
PRIOR FILING DATE: 1999-02-12
NUMBER OF SEQ ID NOS: 1304
SOFTWARE: Patent.pm
SEQ ID NO 572
LENGTH: 499
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: allele
LOCATION: 271
OTHER INFORMATION: 12-215-272 : deletion T
NAME/KEY: misc_binding
LOCATION: 251_270
OTHER INFORMATION: 12-215-272.misl, potential
NAME/KEY: primer_bind
LOCATION: 1..20
OTHER INFORMATION: upstream amplification primer
NAME/KEY: primer_bind
LOCATION: 479..499
OTHER INFORMATION: downstream amplification primer, complement
NAME/KEY: misc_feature
LOCATION: 87
OTHER INFORMATION: n=a, g, c or t
US-09-641-638-572

Query Match 0.8%; Score 18; DB 4; Length 499;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 751 CACAGCAACACAGCAA 768
DB 76 CACAGCAACACAGCAA 59

RESULT 31
US-09-221-017B-390/c
Sequence 390, Application US/09221017B
Patent No. 6444799
GENERAL INFORMATION:
APPLICANT: Ross, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF

NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: PALO ALTO
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221,017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Monroy, Gladys H
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 390:
SEQUENCE CHARACTERISTICS:
LENGTH: 659 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORYPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc feature
LOCATION: 1...659
US-09-221-017B-390

Query Match 0.8%; Score 18; DB 4; Length 659;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1341 TAACTGAGGCTCATTAAG 1358
|||||
Db 525 TAACTGAGGCTCATTAAG 508

RESULT 32
US-09-107-532A-2361
Sequence 2361, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02154
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Arinello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 2361:
SEQUENCE CHARACTERISTICS:
LENGTH: 987 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...987
SEQUENCE DESCRIPTION: SEQ ID NO: 2361:
US-09-107-532A-2361

Query Match 0.8%; Score 18; DB 4; Length 987;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1455 TAGATGATTTTGTGGCAT 1472
|||||
Db 446 TAGATGATTTTGTGGCAT 463

RESULT 33
US-09-107-532A-1110/C
Sequence 1110, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02154
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 1110:
SEQUENCE CHARACTERISTICS:
LENGTH: 1161 base pairs
TYPE: nucleic acid
STRADEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1..1161
SEQUENCE DESCRIPTION: SEQ ID NO: 1110:
US-09-107-532A-1110

Query Match 0.8%; Score 18; DB 4; Length 1161;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 AACGCTGAAGATCAAT 136
|||||
DB 408 AAGCGTGAAGATCAAT 391

RESULT 34
US-08-916-421B-1
Sequence 1, Application US/08916421B
Patent No. 6503729
GENERAL INFORMATION:
APPLICANT: Bult et al.
TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanoco
Patent No. 6503729
TITLE OF INVENTION: jannaschii
FILE REFERENCE: PB275
CURRENT APPLICATION NUMBER: US/08/916,421B
CURRENT FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 60/024,428
PRIOR FILING DATE: 1996-08-22
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 1664976
TYPE: DNA
ORGANISM: Methanococcus jannaschii
FEATURE:
NAME/KEY: misc feature
LOCATION: (28252) .. (28222)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (28257) .. (28258)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (84773) .. (84773)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (84808) .. (84808)
OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc_feature
LOCATION: (84812) .. (84812)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98170) .. (98120)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98159) .. (98159)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98239) .. (98239)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98266) .. (98266)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98343) .. (98343)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (10398) .. (10398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (148948) .. (148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (16385) .. (16385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (191989) .. (191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (191995) .. (191995)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (231980) .. (231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234187) .. (234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234220) .. (234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234814) .. (234814)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (309398) .. (309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (309418) .. (309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (312837) .. (312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (312993) .. (312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (319226) .. (319226)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559167) .. (559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559241) .. (559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (600992) .. (600992)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (622708) .. (622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature

```

LOCATION: (657081)..(657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (674435)..(674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (741684)..(741684)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (855539)..(855539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1130881)..(1130881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1349473)..(1349473)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1349491)..(1349491)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1470091)..(1470091)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1569020)..(1569020)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1602912)..(1602912)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1603734)..(1603734)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1637998)..(1637998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1664854)..(1664854)
OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

```

```

Query Match 0.8%; Score 18; DB 4; Length 1664976;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Oy 2035 TAATGATGATGGCGCAA 2052

Db 789070 TAATGATGATGGCGCAA 789087

RESULT 35

US-08-448-386A-10

Sequence 10, Application US/08448386A

Patent No. 5840708

GENERAL INFORMATION:

APPLICANT: START Technology, Partnership

TITLE OF INVENTION: Administration of Oligonucleotides

TITLE OF INVENTION: Antisense to Dopamine Receptor mRNA for Diagnosis and

TITLE OF INVENTION: Treatment of Pathological Conditions of the Dopaminergic

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dann, Dorfman, Herrell and Skillman

STREET: 1601 Market Street Suite 720

CITY: Philadelphia

STATE: PA

COUNTRY: USA

ZIP: 19103-2307

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/448,386A

FILING DATE: 14-DEC-1993

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/991,582

FILING DATE: 14-DEC-1992

ATTORNEY/AGENT INFORMATION:

NAME: Reed, Janet E.

REGISTRATION NUMBER: 36,252

TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 563-4100

TELEFAX: (215) 563-4044

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA to mRNA

HYPOTHETICAL: NO

ANTI-SENSE: YES

US-08-448-386A-10

Query Match 0.7%; Score 17; DB 2; Length 20;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1735 CCTACTCCTCATCTCCAG 1751

Db 3 CCTACTCCTCATCTCCAG 19

RESULT 36

US-08-816-426-10

Sequence 10, Application US/08816426

Patent No. 6025193

GENERAL INFORMATION:

APPLICANT: START Technology, Partnership

TITLE OF INVENTION: Administration of Oligonucleotides

TITLE OF INVENTION: Antisense to Dopamine Receptor mRNA for Diagnosis and

;; TITLE OF INVENTION: Treatment of Pathological Conditions of the Dopaminergic
;; TITLE OF INVENTION: Nervous System
;; NUMBER OF SEQUENCES: 13
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
;; STREET: 1601 Market Street Suite 720
;; CITY: Philadelphia
;; STATE: PA
;; COUNTRY: USA
;; ZIP: 19103-2307
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/816,426
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/448,386
;; FILING DATE:
;; APPLICATION NUMBER: US 07/991,582
;; FILING DATE: 14-DEC-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Reed, Janet E.
;; REGISTRATION NUMBER: 36,252
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (215) 563-4100
;; TELEFAX: (215) 563-4044
;; INFORMATION FOR SEQ ID NO: 10:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 20 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA to mRNA
;; HYPOTHETICAL: NO
;; ANTI-SENSE: YES
;; US-08-816-426-10

Query Match 0.7%; Score 17; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1735 CCTACCTCATCTCCAG 1751
Db 3 CCTACCTCATCTCCAG 19

RESULT 37
PCT-US93-12161-10
;; Sequence 10, Application PC/TUS9312161
;; GENERAL INFORMATION:
;; APPLICANT: START Technology Partnership
;; TITLE OF INVENTION: Administration of
;; TITLE OF INVENTION: Oligonucleotides Antisense to Dopamine Receptor
;; TITLE OF INVENTION: mRNA for Diagnosis and Treatment of Pathological
;; TITLE OF INVENTION: Conditions of the Dopaminergic Nervous System
;; NUMBER OF SEQUENCES: 13
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Dann, Dorfman, Herrell and
;; STREET: 1601 Market Street, Suite 720
;; CITY: Philadelphia
;; STATE: PA
;; COUNTRY: USA
;; ZIP: 19103-2307
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version

;; SOFTWARE: #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US93/12161
;; FILING DATE: 14-DEC-1993
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/991,582
;; FILING DATE: 14-DEC-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Reed, Janet E.
;; REGISTRATION NUMBER: 36,252
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (215) 563-4100
;; TELEFAX: (215) 563-4044
;; INFORMATION FOR SEQ ID NO: 10:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 20 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA to mRNA
;; HYPOTHETICAL: NO
;; ANTI-SENSE: YES
;; PCT-US93-12161-10

Query Match 0.7%; Score 17; DB 5; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1735 CCTACCTCATCTCCAG 1751
Db 3 CCTACCTCATCTCCAG 19

RESULT 38
US-08-981-462-56
;; Sequence 56, Application US/08981462
;; Patent No. 6054275
;; GENERAL INFORMATION:
;; APPLICANT: Morgan, Una
;; APPLICANT: Thompson, Richard C.A.
;; TITLE OF INVENTION: NOVEL DETECTION METHODS FOR
;; TITLE OF INVENTION: CRYPTOSPORIDIUM
;; NUMBER OF SEQUENCES: 68
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borum
;; STREET: 233 South Wacker Drive/6300 Sears Tower
;; CITY: Chicago
;; STATE: Illinois
;; COUNTRY: United States of America
;; ZIP: 60606-6402
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/981,462
;; FILING DATE:
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: PCT/AU96/00387
;; FILING DATE: 25-JUN-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Cawley, Jr., Thomas A.
;; REGISTRATION NUMBER: 40,944
;; REFERENCE/DOCKET NUMBER: 28594/34423
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (312) 474-6300
;; TELEFAX: (312) 474-0448
;; INFORMATION FOR SEQ ID NO: 56:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 24 base pairs

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-981-462-56

Query Match 0.7%; Score 17; DB 3; Length 24;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2090 GAAGTCCAATAGAA 2106
DB 2 GAAGTCCAATAGAA 18

RESULT 39
US-09-491-356C-15
Sequence 15, Application US/09491356C
Patent No. 6566061
GENERAL INFORMATION:
APPLICANT: Philibert, Robert A.
APPLICANT: Delisi, Lynn
TITLE OF INVENTION: IDENTIFICATION OF POLYMORPHISMS IN THE PCTG4 REGION OF XQ13
FILE REFERENCE: 9465.6US11
CURRENT APPLICATION NUMBER: US/09/491,356C
CURRENT FILING DATE: 2000-01-26
PRIOR APPLICATION NUMBER: PCT/US99/09365
PRIOR FILING DATE: 1999-04-29
PRIOR APPLICATION NUMBER: 60/083,465
PRIOR FILING DATE: 1998-04-29
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn version 3.1
SEQ ID NO 15
LENGTH: 238
TYPE: DNA
ORGANISM: Homo sapiens
US-09-491-356C-15

Query Match 0.7%; Score 17; DB 4; Length 238;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 755 AGCAACACAGCAACAC 771
DB 141 AGCAACACAGCAACAC 157

RESULT 40
US-09-491-356C-13
Sequence 13, Application US/09491356C
Patent No. 6566061
GENERAL INFORMATION:
APPLICANT: Philibert, Robert A.
APPLICANT: Delisi, Lynn
TITLE OF INVENTION: IDENTIFICATION OF POLYMORPHISMS IN THE PCTG4 REGION OF XQ13
FILE REFERENCE: 9465.6US11
CURRENT APPLICATION NUMBER: US/09/491,356C
CURRENT FILING DATE: 2000-01-26
PRIOR APPLICATION NUMBER: PCT/US99/09365
PRIOR FILING DATE: 1999-04-29
PRIOR APPLICATION NUMBER: 60/083,465
PRIOR FILING DATE: 1998-04-29
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn version 3.1
SEQ ID NO 13
LENGTH: 253
TYPE: DNA
ORGANISM: Homo sapiens
US-09-491-356C-13

Query Match 0.7%; Score 17; DB 4; Length 253;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 755 AGCAACACAGCAACAC 771
DB 156 AGCAACACAGCAACAC 172

RESULT 41
US-09-491-356C-14
Sequence 14, Application US/09491356C
Patent No. 6566061
GENERAL INFORMATION:
APPLICANT: Philibert, Robert A.
APPLICANT: Delisi, Lynn
TITLE OF INVENTION: IDENTIFICATION OF POLYMORPHISMS IN THE PCTG4 REGION OF XQ13
FILE REFERENCE: 9465.6US11
CURRENT APPLICATION NUMBER: US/09/491,356C
CURRENT FILING DATE: 2000-01-26
PRIOR APPLICATION NUMBER: PCT/US99/09365
PRIOR FILING DATE: 1999-04-29
PRIOR APPLICATION NUMBER: 60/083,465
PRIOR FILING DATE: 1998-04-29
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn version 3.1
SEQ ID NO 14
LENGTH: 265
TYPE: DNA
ORGANISM: Homo sapiens
US-09-491-356C-14

Query Match 0.7%; Score 17; DB 4; Length 265;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 755 AGCAACACAGCAACAC 771
DB 156 AGCAACACAGCAACAC 172

RESULT 42
US-09-491-356C-16
Sequence 16, Application US/09491356C
Patent No. 6566061
GENERAL INFORMATION:
APPLICANT: Philibert, Robert A.
APPLICANT: Delisi, Lynn
TITLE OF INVENTION: IDENTIFICATION OF POLYMORPHISMS IN THE PCTG4 REGION OF XQ13
FILE REFERENCE: 9465.6US11
CURRENT APPLICATION NUMBER: US/09/491,356C
CURRENT FILING DATE: 2000-01-26
PRIOR APPLICATION NUMBER: PCT/US99/09365
PRIOR FILING DATE: 1999-04-29
PRIOR APPLICATION NUMBER: 60/083,465
PRIOR FILING DATE: 1998-04-29
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn version 3.1
SEQ ID NO 16
LENGTH: 265
TYPE: DNA
ORGANISM: Homo sapiens
US-09-491-356C-16

Query Match 0.7%; Score 17; DB 4; Length 265;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 755 AGCAACACAGCAACAC 771
DB 156 AGCAACACAGCAACAC 172

```
RESULT 43
US-09-491-356C-17
; Sequence 17, Application US/09491356C
; Patent No. 6566061
; GENERAL INFORMATION:
; APPLICANT: Philibert, Robert A.
; APPLICANT: Gims, Edward I.
; APPLICANT: Delisi, Lynn
; TITLE OF INVENTION: IDENTIFICATION OF POLYMORPHISMS IN THE PCTG4 REGION OF XQ13
; FILE REFERENCE: 9465.6US11
; CURRENT APPLICATION NUMBER: US/09/491,356C
; CURRENT FILING DATE: 2000-01-26
; PRIOR APPLICATION NUMBER: PCT/US99/09365
; PRIOR FILING DATE: 1999-04-29
; PRIOR APPLICATION NUMBER: 60/083,465
; PRIOR FILING DATE: 1998-04-29
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 17
; LENGTH: 265
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-491-356C-17

Query Match      0.7%; Score 17; DB 4; Length 265;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      755 AGCAACAAGACGACAC 771
DB      156 AGCAACAAGACGACAC 172

RESULT 44
US-09-491-356C-18
; Sequence 18, Application US/09491356C
; Patent No. 6566061
; GENERAL INFORMATION:
; APPLICANT: Philibert, Robert A.
; APPLICANT: Gims, Edward I.
; APPLICANT: Delisi, Lynn
; TITLE OF INVENTION: IDENTIFICATION OF POLYMORPHISMS IN THE PCTG4 REGION OF XQ13
; FILE REFERENCE: 9465.6US11
; CURRENT APPLICATION NUMBER: US/09/491,356C
; CURRENT FILING DATE: 2000-01-26
; PRIOR APPLICATION NUMBER: PCT/US99/09365
; PRIOR FILING DATE: 1999-04-29
; PRIOR APPLICATION NUMBER: 60/083,465
; PRIOR FILING DATE: 1998-04-29
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 18
; LENGTH: 265
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-491-356C-18

Query Match      0.7%; Score 17; DB 4; Length 265;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      755 AGCAACAAGACGACAC 771
DB      156 AGCAACAAGACGACAC 172

RESULT 45
US-08-175-388-2/C
; Sequence 2, Application US/08175388
; Patent No. 5641674
; GENERAL INFORMATION:
; APPLICANT: DEHOUX, PIERRE
```

```
APPLICANT: DAVIES, JULIAN
; TITLE OF INVENTION: Nucleotide Sequences and Proteins
; TITLE OF INVENTION: Confering Cycloheximide Resistance
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/175,388
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR92/00685
; FILING DATE: 15-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9108906
; FILING DATE: 15-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5641674man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 660-071-0 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 321 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-175-388-2

Query Match      0.7%; Score 17; DB 1; Length 321;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1922 AAATTGCTGTGTTGA 1938
DB      172 AAATTGCTGTGTTGA 156

RESULT 46
US-08-779-620-2/C
; Sequence 2, Application US/08779620
; Patent No. 5830997
; GENERAL INFORMATION:
; APPLICANT: DEHOUX, PIERRE
; APPLICANT: DAVIES, JULIAN
; TITLE OF INVENTION: Nucleotide Sequences and Proteins
; TITLE OF INVENTION: Confering Cycloheximide Resistance
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/779,620
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175,388
FILING DATE: 09-JUN-1994
APPLICATION NUMBER: PCT/FR92/00685
FILING DATE: 15-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 9108906
FILING DATE: 15-JUL-1991
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, No. 583097/man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 660-071-0 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 321 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-779-620-2

Query Match 0.7%; Score 17; DB 2; Length 321;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1922 AAATTGCTGTTGA 1938
Db 172 AAATTGCTGTTGA 156

RESULT 47
US-08-818-726-2/c
Sequence 2, Application US/08818726
Patent No. 5958687
GENERAL INFORMATION:
APPLICANT: DEHOX, PIERRE
APPLICANT: DAVIES, JULIAN
TITLE OF INVENTION: Nucleotide Sequences and Proteins
TITLE OF INVENTION: Confering Cycloheximide Resistance
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,
ADDRESS: P.C.
STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,726
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175,388
FILING DATE: 09-JUN-1994
APPLICATION NUMBER: PCT/FR92/00685
FILING DATE: 15-JUL-1992

PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 9108906
FILING DATE: 15-JUL-1991
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, No. 5958687/man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 660-071-0 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 321 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-818-726-2

Query Match 0.7%; Score 17; DB 2; Length 321;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1922 AAATTGCTGTTGA 1938
Db 172 AAATTGCTGTTGA 156

RESULT 48
US-08-961-083-149
Sequence 149, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Chol et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 149:
SEQUENCE CHARACTERISTICS:
LENGTH: 322 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-083-149

Query Match 0.7%; Score 17; DB 3; Length 322;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1439 AATAAAGAAAATTGCT 1455
|||||
38 AATAAAGAAAATTGCT 54

RESULT 49
US-09-536-784-149
Sequence 149, Application US/09536784
Patent No. 6573082

GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSER: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/536,784
FILING DATE: 30-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: OCT-30-1997

ATTORNEY/AGENT INFORMATION:

NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB340P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 149:

SEQUENCE CHARACTERISTICS:
LENGTH: 322 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

US-09-536-784-149
SEQUENCE DESCRIPTION: SEQ ID NO: 149:

Query Match 0.7%; Score 17; DB 4; Length 322;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1439 AATAAAGAAAATTGCT 1455
|||||
38 AATAAAGAAAATTGCT 54

RESULT 50
US-09-056-556-216/C

Sequence 216, Application US/09056556
Patent No. 6350456

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESSER: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle

TREATM

STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 216:

SEQUENCE CHARACTERISTICS:

LENGTH: 470 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA

US-09-056-556-216

Query Match 0.7%; Score 17; DB 4; Length 470;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2138 ACACCTGCTGAGCCGAGA 2154
|||||
DB 159 ACACCTGCTGAGCCGAGA 143

RESULT 51

US-09-072-596-211/C

Sequence 211, Application US/09072596
Patent No. 6458366

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
APPLICANT: Campos-Neto, Antonia
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.

APPLICANT: Hendrickson, Ronald C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSER: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9

```
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 211:
SEQUENCE CHARACTERISTICS:
LENGTH: 470 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-09-072-586-211

Query Match      0.7%; Score 17; DB 4; Length 470;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      2138 ACACCTGCTGAGCCAGA 2154
Db      159 ACACCTGCTGAGCCAGA 143

RESULT 52
US-09-134-001C-1953/C
Sequence 1953, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT FILING DATE: 1998-08-13
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 1953
LENGTH: 633
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1953

Query Match      0.7%; Score 17; DB 4; Length 633;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1610 ATATCTAGTGATGAAG 1626
Db      147 ATATCTAGTGATGAAG 131

RESULT 53
US-09-221-017B-1007/C
Sequence 1007, Application US/09221017B
Patent No. 6444799
GENERAL INFORMATION:
APPLICANT: ROSS, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FORSTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
```

```
APPLICATION NUMBER: US/09/221,017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Monroy, Gladys H
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 1007:
SEQUENCE CHARACTERISTICS:
LENGTH: 658 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORYPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1...658
US-09-221-017B-1007

Query Match      0.7%; Score 17; DB 4; Length 658;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      703 AACAACTGGGTACCTT 719
Db      361 AACAACTGGGTACCTT 345

RESULT 54
US-09-328-352-1650
Sequence 1650, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT FILING DATE: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 1650
LENGTH: 687
TYPE: DNA
ORGANISM: Acinetobacter baumannii
US-09-328-352-1650

Query Match      0.7%; Score 17; DB 4; Length 687;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1862 ACAGTTGAGTTAAAAA 1878
Db      370 ACAGTTGAGTTAAAAA 386
```

RESULT 55
US-08-981-462-1/C
; Sequence 1, Application US/08981462
; Patent No. 6054275
; GENERAL INFORMATION:
; APPLICANT: Morgan, Una
; APPLICANT: Thompson, Richard C.A.
; TITLE OF INVENTION: NOVEL DETECTION METHODS FOR
; TITLE OF INVENTION: CRYPTOSPORIDIUM
; NUMBER OF SEQUENCES: 68
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive/6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/981,462
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/00387
; FILING DATE: 25-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Cawley, Jr., Thomas A.
; REGISTRATION NUMBER: 40,944
; REFERENCE/DOCKET NUMBER: 28594/34423
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 688 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-981-462-1

Query Match 0.7%; Score 17; DB 3; Length 688;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2090 GAAGATCCCAATTAAGA 2106
DB 498 GAAGATCCCAATTAAGA 482

RESULT 56
US-09-470-191-60/C
; Sequence 60, Application US/09470191
; Patent No. 6465633
; GENERAL INFORMATION:
; APPLICANT: Skelky, Yasir
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods of Their Use in
; TITLE OF INVENTION: the Treatment, Prevention and Diagnosis of Tuberculosis
; FILE REFERENCE: 014058-008910US
; CURRENT APPLICATION NUMBER: US/09/470,191
; CURRENT FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: US 60/113,952
; PRIOR FILING DATE: 1998-12-24
; NUMBER OF SEQ ID NOS: 97
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 60

LENGTH: 706
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (1)...(706)
; OTHER INFORMATION: n = any nucleotide
US-09-470-191-60

Query Match 0.7%; Score 17; DB 4; Length 706;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 422 CAACATCGTAGGTGG 438
DB 101 CAACATCGTAGGTGG 85

RESULT 57
US-09-071-035-373
; Sequence 373, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gai H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 373:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 762 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-09-071-035-373

Query Match 0.7%; Score 17; DB 4; Length 762;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 789 AAAGTAATGACATTGAT 805
DB 16 AAAGTAATGACATTGAT 32

RESULT 58
US-09-501-558-1
; Sequence 1, Application US/09501558
; Patent No. 6403784

```
; GENERAL INFORMATION:
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Mathur, Brian
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6403784e1 Human Uncoupling Proteins and
; FILE REFERENCE: Lex-0012-USA
; CURRENT APPLICATION NUMBER: US/09/501,558
; CURRENT FILING DATE: 2000-02-09
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 876
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-501-558-1

Query Match          0.7%; Score 17; DB 4; Length 876;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1025 CCAGATTCAGGCCAGA 1041
      |||||
Db      99  CCAGATTCAGGCCAGA 115

RESULT 59
US-09-501-558-3
; Sequence 3, Application US/09501558
; Patent No. 6403784
; GENERAL INFORMATION:
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Mathur, Brian
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6403784e1 Human Uncoupling Proteins and
; FILE REFERENCE: Lex-0012-USA
; CURRENT APPLICATION NUMBER: US/09/501,558
; CURRENT FILING DATE: 2000-02-09
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 882
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-501-558-3

Query Match          0.7%; Score 17; DB 4; Length 882;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1025 CCAGATTCAGGCCAGA 1041
      |||||
Db      99  CCAGATTCAGGCCAGA 115

RESULT 60
US-09-252-991A-1381
; Sequence 1381, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
```

```
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1381
; LENGTH: 909
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1381

Query Match          0.7%; Score 17; DB 4; Length 909;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      905 AGAGGTGTCAGTGCC 921
      |||||
Db      268 AGAGGTGTCAGTGCC 284

RESULT 61
US-09-328-352-3547
; Sequence 3547, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 3547
; LENGTH: 909
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-3547

Query Match          0.7%; Score 17; DB 4; Length 909;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1216 TGAACGTGTAAATC 1232
      |||||
Db      213 TGAACGTGTAAATC 229

RESULT 62
US-09-107-532A-2698
; Sequence 2698, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
```


NAME: Ariniello, Pamela Deneke
 REGISTRATION NUMBER: 40,489
 REFERENCE/DOCKET NUMBER: GTC-012
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (781)893-5007
 TELEFAX: (781)893-8277
 INFORMATION FOR SEQ ID NO: 2698:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 975 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: circular
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Enterococcus faecium
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (b) LOCATION 1...975
 SEQUENCE DESCRIPTION: SEQ ID NO: 2698:
 US-09-107-532A-2698

Query Match 0.7%; Score 17; DB 4; Length 975;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1442 AAGAAATTTGTTAG 1458
 DB 949 AAGAAATTTGTTAG 965

RESULT 63
 US-08-839-711-6
 Sequence 6, Application US/08839711
 Patent No. 6033870
 GENERAL INFORMATION:
 APPLICANT: Hillman, Jennifer L.
 TITLE OF INVENTION: NOVEL HUMAN TRANSMEMBRANE
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Incyte Pharmaceuticals, Inc.
 STREET: 3174 Porter Drive
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94304
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/839,711
 FILING DATE: Filed Herewith
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Billings, Lucy J.
 REGISTRATION NUMBER: 36,749
 REFERENCE/DOCKET NUMBER: PF-0270 US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-855-0555
 TELEFAX: 415-845-4166
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1077 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear

IMMEDIATE SOURCE:
 LIBRARY: GenBank
 CLONE: 476343
 US-08-839-711-6

Query Match 0.7%; Score 17; DB 3; Length 1077;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1275 TAACGTCTAAAGAA 1291
 DB 691 TAACGTCTAAAGAA 707

RESULT 64
 US-09-328-352-3376
 Sequence 3376, Application US/09328352
 Patent No. 6562958
 GENERAL INFORMATION:
 APPLICANT: Gary L. Breton et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 FILE REFERENCE: GTC99-03PA
 CURRENT APPLICATION NUMBER: US/09/328,352
 CURRENT FILING DATE: 1999-06-04
 NUMBER OF SEQ ID NOS: 8252
 SEQ ID NO 3376
 LENGTH: 1080
 TYPE: DNA
 ORGANISM: Acinetobacter baumannii
 US-09-328-352-3376

Query Match 0.7%; Score 17; DB 4; Length 1080;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1546 TCGTATGCTCATTTAG 1562
 DB 597 TCGTATGCTCATTTAG 613

RESULT 65
 US-09-252-991A-1287
 Sequence 1287, Application US/09252991A
 Patent No. 6551795
 GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 CURRENT FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 1287
 LENGTH: 1092
 TYPE: DNA
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-1287

Query Match 0.7%; Score 17; DB 4; Length 1092;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 905 AGAGGTGTCAGTCC 921
 DB 282 AGAGGTGTCAGTCC 298

RESULT 66

US-09-252-991A-1184/c
; Sequence 1184, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1184
; LENGTH: 1230
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1184

Query Match 0.7%; Score 17; DB 4; Length 1230;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 905 AGAGGTGTCAGTSCC 921
|||
Db 1126 AGAGGTGTCAGTSCC 1110

RESULT 67
US-09-359-301A-9/c
; Sequence 9, Application US/09359301A
; Patent No. 6426185
; GENERAL INFORMATION:
; APPLICANT: Kumagai, Monto H.
; APPLICANT: della-Cioppa, Guy R.
; APPLICANT: Erwin, Robert L.
; APPLICANT: McGee, David R.
; TITLE OF INVENTION: METHOD OF DETERMINING THE PRESENCE OF A
; TITLE OF INVENTION: TRAIT IN A PLANT BY TRANSFECTING A NUCLEIC ACID SEQUENCE OF
; TITLE OF INVENTION: A DONOR PLANT INTO A DIFFERENT HOST PLANT IN AN ANTI-SENSE
; FILE REFERENCE: 008010137US04
; CURRENT APPLICATION NUMBER: US/09/359,301A
; CURRENT FILING DATE: 1999-07-21
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 1543
; TYPE: DNA
; ORGANISM: Ribgrass mosaic virus
US-09-359-301A-9

Query Match 0.7%; Score 17; DB 4; Length 1543;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 94 GAATTGACTCCTGATG 110
|||
Db 1213 GAATTGACTCCTGATG 1197

RESULT 68
US-09-107-532A-2089
; Sequence 2089, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A. Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Arinello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 2089:
SEQUENCE CHARACTERISTICS:
LENGTH: 1578 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...1578
SEQUENCE DESCRIPTION: SEQ ID NO: 2089:
US-09-107-532A-2089

Query Match 0.7%; Score 17; DB 4; Length 1578;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1442 AAAGAAAATTGTAGA 1458
|||
Db 134 AAAGAAAATTGTAGA 150

RESULT 69
US-09-613-303-3
; Sequence 3, Application US/09613303
; Patent No. 6495347
; GENERAL INFORMATION:
; APPLICANT: Siegel, Marvin
; APPLICANT: Chu, N. Randall
; APPLICANT: Mizzen, Lee A.
; TITLE OF INVENTION: INDUCTION OF A TH1-LIKE RESPONSE IN VITRO
; FILE REFERENCE: 12071/002001
; CURRENT APPLICATION NUMBER: US/09/613,303
; CURRENT FILING DATE: 2000-07-10
; PRIOR APPLICATION NUMBER: US 60/143,757
; PRIOR FILING DATE: 1999-07-08
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1623
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:

OTHER INFORMATION: fusion sequence
NAME/KEY: CDS
LOCATION: (1)...(1620)
US-09-613-303-3

Query Match 0.7%; Score 17; DB 4; Length 1623;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 422 CAACATCGTAGAGTGG 438
DB 1311 CAACATCGTAGAGTGG 1327

RESULT 70
US-08-487-596-3/C
Sequence 3, Application US/08487596
Patent No. 6440681
GENERAL INFORMATION:
APPLICANT: Elliot, Kathryn J.
APPLICANT: Ellis, Steven B.
APPLICANT: Harpold, Michael M.
TITLE OF INVENTION: METHODS FOR IDENTIFYING AGONISTS AND
TITLE OF INVENTION: ANTIGONISTS FOR HUMAN NEURONAL
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,596
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO US94/02447
FILING DATE: 08-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/149,503
FILING DATE: 08-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/028,031
FILING DATE: 08-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/938,154
FILING DATE: 30-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/504,455
FILING DATE: 03-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L.
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-9951
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-238-0999
TELEFAX: 619-238-0062
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1654 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS

LOCATION: 39..1553
OTHER INFORMATION: /product= "ALPHA-3 SUBUNIT"
US-08-487-596-3

Query Match 0.7%; Score 17; DB 4; Length 1654;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1615 CAGTGATGAGGAGATG 1631
DB 813 CAGTGATGAGGAGATG 797

RESULT 71
US-08-466-589-3/C
Sequence 3, Application US/08466589
Patent No. 5857489
GENERAL INFORMATION:
APPLICANT: Elliot, Kathryn J.
APPLICANT: Ellis, Steven B.
APPLICANT: Harpold, Michael M.
TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE
TITLE OF INVENTION: RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,589
FILING DATE: June 5, 1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/028,031
FILING DATE: March 8, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L.
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-9950
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-238-0999
TELEFAX: 619-238-0062
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1756 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 39..1553
US-08-466-589-3

Query Match 0.7%; Score 17; DB 2; Length 1756;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1615 CAGTGATGAGGAGATG 1631
DB 813 CAGTGATGAGGAGATG 797

RESULT 72

US-08-700-636-3/c
; Sequence 3, Application US/08700636
; Patent No. 5910582
; GENERAL INFORMATION:
; APPLICANT: Elliot, Kathryn J.
; APPLICANT: Ellis, Steven B.
; APPLICANT: Harpold, Michael M.
; TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE
; TITLE OF INVENTION: RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/700.636
; FILING DATE: 16-JUL-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/028.031
; FILING DATE: 08-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31.192
; REFERENCE/DOCKET NUMBER: P41 9368
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-546-4737
; TELEFAX: 619-546-9392
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1756 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 39..1553
; US-08-700-636-3

Query Match 0.7%; Score 17; DB 2; Length 1756;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1615 CAGTGATGAGAGATG 1631
|||
Db 813 CAGTGATGAGAGATG 797

RESULT 73
US-08-467-574-3/c
; Sequence 3, Application US/08467574
; Patent No. 6022704
; GENERAL INFORMATION:
; APPLICANT: Elliot, Kathryn J.
; APPLICANT: Ellis, Steven B.
; APPLICANT: Harpold, Michael M.
; TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE
; TITLE OF INVENTION: RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClaim
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA

COUNTRY: USA
ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,574
FILING DATE: June 5, 1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/028,031
FILING DATE: March 8, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-9949
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-238-0999
TELEFAX: 619-238-0062
TELFX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1756 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 39..1553
US-08-467-574-3

Query Match 0.7%; Score 17; DB 3; Length 1756;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1615 CAGTGATGAGAGATG 1631
|||
Db 813 CAGTGATGAGAGATG 797

RESULT 74
US-09-217-345-3/c
; Sequence 3, Application US/09217345
; Patent No. 6303753
; GENERAL INFORMATION:
; APPLICANT: Elliot, Kathryn J.
; APPLICANT: Ellis, Steven B.
; APPLICANT: Harpold, Michael M.
; TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE
; TITLE OF INVENTION: RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Ertman White & McAniff
; STREET: 4250 Executive Square, 7th Floor
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/217,345
; FILING DATE: 21-DEC-98
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/467,574
; FILING DATE: 05-JUN-95
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/466,589,
FILING DATE: 05-JUN-95
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/028,031
FILING DATE: 08-MAR-93
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 24735-9949B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-450-8400
TELEFAX: 619-587-5360
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1756 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 39..1553
US-09-217-345-3

Query Match 0.7%; Score 17; DB 4; Length 1756;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1615 CAGTGATGAAGGATG 1631
DB 813 CAGTGATGAAGGATG 797

RESULT 75
US-08-755-587-1
Sequence 1, Application US/08755587
Patent No. 6045997
GENERAL INFORMATION:
APPLICANT: Futreal, Phillip A
APPLICANT: Wooster, Richard F
APPLICANT: Ashworth, Alan
APPLICANT: Stratton, Michael R
TITLE OF INVENTION: Materials and methods relating to the
TITLE OF INVENTION: identification and sequencing of the BRCA2 cancer
TITLE OF INVENTION: susceptibility gene and uses thereof.
NUMBER OF SEQUENCES: 222
CORRESPONDENCE ADDRESS:
ADDRESSER: Bell Seltzer Park & Gibson
STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
CITY: Raleigh
STATE: NC
COUNTRY: USA
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/755,587
FILING DATE: 25-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9523959.6
FILING DATE: 23-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9525555.0
FILING DATE: 14-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9617961.9
FILING DATE: 28-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kenneth D Sibley
REGISTRATION NUMBER: 31,665

REFERENCE/DOCKET NUMBER: 5405-135
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1917 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-755-587-1

Query Match 0.7%; Score 17; DB 3; Length 1917;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 652 TCTGCAATTCAAGAA 668
DB 582 TCTGCAATTCAAGAA 598

RESULT 76
US-09-613-303-16
Sequence 16, Application US/09613303
Patent No. 6495347
GENERAL INFORMATION:
APPLICANT: Siegel, Marvin
APPLICANT: Chu, N. Randall
APPLICANT: Mizzen, Lee A
TITLE OF INVENTION: INDUCTION OF A TH1-LIKE RESPONSE IN VITRO
FILE REFERENCE: 12071/002001
CURRENT APPLICATION NUMBER: US/09/613,303
CURRENT FILING DATE: 2000-07-10
PRIOR APPLICATION NUMBER: US 60/143,757
PRIOR FILING DATE: 1999-07-08
NUMBER OF SEQ ID NOS: 55
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 16
LENGTH: 1920
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(1917)
US-09-613-303-16

Query Match 0.7%; Score 17; DB 4; Length 1920;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 422 CAACATCGTGAAGGTGG 438
DB 1311 CAACATCGTGAAGGTGG 1327

RESULT 77
US-09-613-303-28
Sequence 28, Application US/09613303
Patent No. 6495347
GENERAL INFORMATION:
APPLICANT: Siegel, Marvin
APPLICANT: Chu, N. Randall
APPLICANT: Mizzen, Lee A
TITLE OF INVENTION: INDUCTION OF A TH1-LIKE RESPONSE IN VITRO
FILE REFERENCE: 12071/002001
CURRENT APPLICATION NUMBER: US/09/613,303
CURRENT FILING DATE: 2000-07-10
PRIOR APPLICATION NUMBER: US 60/143,757
PRIOR FILING DATE: 1999-07-08
NUMBER OF SEQ ID NOS: 55
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 28
LENGTH: 1947
TYPE: DNA

```
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: fusion sequence
NAME/KEY: CDS
LOCATION: (1)...(1944)
US-09-613-303-28
```

```
Query Match      0.7%; Score 17; DB 4; Length 1947;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      422 CAACATCGTGAAGTGG 438
          |||||
Db      1635 CAACATCGTGAAGTGG 1651
```

```
RESULT 78
US-09-370-807-11
Sequence 11, Application US/09370807
Patent No. 6297034
GENERAL INFORMATION:
APPLICANT: Cahoon, Rebecca E.
APPLICANT: Falco, S. Carl
APPLICANT: Rafalski, J. Antoni
APPLICANT: Sakai, Hajime
TITLE OF INVENTION: N-End Rule Pathway Enzymes
FILE REFERENCE: BB-1199
CURRENT APPLICATION NUMBER: US/09/370,807
CURRENT FILING DATE: 1999-08-09
EARLIER APPLICATION NUMBER: 60/096,225
EARLIER FILING DATE: August 12, 1998
NUMBER OF SEQ ID NOS: 16
SOFTWARE: Microsoft Office 97
SEQ ID NO 11
LENGTH: 2018
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
NAME/KEY: unsure
LOCATION: (17)
FEATURE:
NAME/KEY: unsure
LOCATION: (19)
FEATURE:
NAME/KEY: unsure
LOCATION: (46)
FEATURE:
NAME/KEY: unsure
LOCATION: (50)
FEATURE:
NAME/KEY: unsure
LOCATION: (80)...(81)
FEATURE:
NAME/KEY: unsure
LOCATION: (113)
FEATURE:
NAME/KEY: unsure
LOCATION: (127)
FEATURE:
NAME/KEY: unsure
LOCATION: (158)
FEATURE:
NAME/KEY: unsure
LOCATION: (190)
FEATURE:
NAME/KEY: unsure
LOCATION: (302)
FEATURE:
NAME/KEY: unsure
LOCATION: (308)
FEATURE:
NAME/KEY: unsure
LOCATION: (324)
```

```
FEATURE:
NAME/KEY: unsure
LOCATION: (346)
FEATURE:
NAME/KEY: unsure
LOCATION: (364)
FEATURE:
NAME/KEY: unsure
LOCATION: (425)
FEATURE:
NAME/KEY: unsure
LOCATION: (435)
FEATURE:
NAME/KEY: unsure
LOCATION: (442)
US-09-370-807-11
```

```
Query Match      0.7%; Score 17; DB 3; Length 2018;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1284 AAAAAGAAATGCTGCT 1300
          |||||
Db      754 AAAAAGAAATGCTGCT 770
```

```
RESULT 79
US-09-921-259-11
Sequence 11, Application US/09921259
Patent No. 6465234
GENERAL INFORMATION:
APPLICANT: Cahoon, Rebecca E.
APPLICANT: Falco, S. Carl
APPLICANT: Rafalski, J. Antoni
APPLICANT: Sakai, Hajime
TITLE OF INVENTION: N-End Rule Pathway Enzymes
FILE REFERENCE: BB-1199
CURRENT APPLICATION NUMBER: US/09/921,259
CURRENT FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: 60/096,225
PRIOR FILING DATE: August 12, 1998
NUMBER OF SEQ ID NOS: 16
SOFTWARE: Microsoft Office 97
SEQ ID NO 11
LENGTH: 2018
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
NAME/KEY: unsure
LOCATION: (17)
NAME/KEY: unsure
LOCATION: (19)
NAME/KEY: unsure
LOCATION: (46)
NAME/KEY: unsure
LOCATION: (50)
NAME/KEY: unsure
LOCATION: (80)...(81)
NAME/KEY: unsure
LOCATION: (113)
NAME/KEY: unsure
LOCATION: (127)
NAME/KEY: unsure
LOCATION: (158)
NAME/KEY: unsure
LOCATION: (190)
NAME/KEY: unsure
LOCATION: (302)
NAME/KEY: unsure
LOCATION: (308)
NAME/KEY: unsure
LOCATION: (324)
NAME/KEY: unsure
```

LOCATION: (348)
NAME/KEY: unsure.
LOCATION: (364)
NAME/KEY: unsure
LOCATION: (425)
NAME/KEY: unsure
LOCATION: (435)
NAME/KEY: unsure
LOCATION: (442)
US-09-921-259-11

Query Match 0.7%; Score 17; DB 4; Length 2018;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1284 AAAAGAAATGTGCT 1300
DB 754 AAAAGAAATGTGCT 770

RESULT 80

US-08-714-918-105/c
Sequence 105 Application US/08714918
Patent No. 6037123

GENERAL INFORMATION:
APPLICANT: Benton, Bret
APPLICANT: Lee, Ving
APPLICANT: Malouin, Francois
APPLICANT: Martin, Patrick K.
APPLICANT: Schmid, Molly B.
APPLICANT: Sun, Dongxu
TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS ANTIBACTERIAL
TITLE OF INVENTION: TARGET GENES
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/714,918
FILING DATE: September 13, 1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/009,102
FILING DATE: December 22, 1995
APPLICATION NUMBER: 60/003,798
FILING DATE: September 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 222/005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510

INFORMATION FOR SEQ ID NO: 105:
SEQUENCE CHARACTERISTICS:
LENGTH: 2255 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-714-918-105

Query Match 0.7%; Score 17; DB 3; Length 2255;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1112 TCAATTCTCTTTGCT 1128
DB 583 TCAATTCTCTTTGCT 567

RESULT 81

US-09-265-315-105/c
Sequence 105 Application US/09265315
Patent No. 6187541

GENERAL INFORMATION:
APPLICANT: Benton, Bret
APPLICANT: Lee, Ving J.
APPLICANT: Malouin, Francois
APPLICANT: Martin, Patrick K.
APPLICANT: Schmid, Molly B.
APPLICANT: Sun, Dongxu
TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/265,315
FILING DATE: March 9, 1999
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/714,918
FILING DATE: September 13, 1996
APPLICATION NUMBER: 60/009,102
FILING DATE: December 22, 1995
APPLICATION NUMBER: 60/003,798
FILING DATE: September 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 240/247
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510

INFORMATION FOR SEQ ID NO: 105:
SEQUENCE CHARACTERISTICS:
LENGTH: 2255 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

US-09-265-315-105

Query Match 0.7%; Score 17; DB 3; Length 2255;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1112 TCAATTCTCTTTGCT 1128
DB 583 TCAATTCTCTTTGCT 567

RESULT 82
US-09-265-315-105/c
Sequence 105, Application US/09265315
Patent No. 6187541
GENERAL INFORMATION:
APPLICANT: Benton, Bret
APPLICANT: Lee, Ving J.
APPLICANT: Malouin, Francois
APPLICANT: Martin, Patrick K.
APPLICANT: Schmid, Molly B.
APPLICANT: Sun, Dongxu
TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/265,315
FILING DATE: March 9, 1999
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/714,918
FILING DATE: September 13, 1996
APPLICATION NUMBER: 60/009,102
FILING DATE: December 22, 1995
APPLICATION NUMBER: 60/003,798
FILING DATE: September 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 240/247
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 105:
SEQUENCE CHARACTERISTICS:
LENGTH: 2255 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-265-315-105

Query Match 0.7%; Score 17; DB 3; Length 2255;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1112 TCAATTCTTCTTGT 1128
|||||
Db 583 TCAATTCTTCTTGT 567

RESULT 83
US-09-266-417-105/c
Sequence 105, Application US/09266417
Patent No. 6228588
GENERAL INFORMATION:
APPLICANT: Benton, Bret
APPLICANT: Lee, Ving J.

APPLICANT: Malouin, Francois
APPLICANT: Martin, Patrick K.
APPLICANT: Schmid, Molly B.
APPLICANT: Sun, Dongxu
TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/266,417
FILING DATE: March 9, 1999
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/714,918
FILING DATE: September 13, 1996
APPLICATION NUMBER: 60/009,102
FILING DATE: December 22, 1995
APPLICATION NUMBER: 60/003,798
FILING DATE: September 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 240/248
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 105:
SEQUENCE CHARACTERISTICS:
LENGTH: 2255 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-266-417-105

Query Match 0.7%; Score 17; DB 3; Length 2255;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1112 TCAATTCTTCTTGT 1128
|||||
Db 583 TCAATTCTTCTTGT 567

RESULT 84
US-08-913-159-12/c
Sequence 12, Application US/08913159
Patent No. 6300109
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Plasmid-derived type II
TITLE OF INVENTION: restriction-modification systems from Lactococcus lactis
NUMBER OF SEQUENCES: 14
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:


```
APPLICATION NUMBER: US/08/913,159
FILING DATE:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: DK 0179/95
FILING DATE: 17-FEB-1995
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 2355 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Lactococcus lactis subsp. cremoris
STRAIN: W39
FEATURE:
NAME/KEY: CDS
LOCATION: 744..1283
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=744
OTHER INFORMATION: /product="L1adII restriction endonuclease"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="ORF"
OTHER INFORMATION: /number=1
OTHER INFORMATION: /standard_name="Gene coding for R.L1adII"
OTHER INFORMATION: /note="The first ten amino acids in this sequence may be
OTHER INFORMATION: doubtful. However, from base 773 this reading frame gives a h
FEATURE:
NAME/KEY: CDS
LOCATION: 1392..2342
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=1392
OTHER INFORMATION: /product="L1adII methylase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="ORF"
OTHER INFORMATION: /number=2
OTHER INFORMATION: /standard_name="Gene coding for M.L1adII"
OTHER INFORMATION: /label="m-l1adII"
OTHER INFORMATION: /note="The sequence shows 60 % identity and 76 % similarity
OTHER INFORMATION: with the Bsp61 methylase."
US-08-913-159-12
Query Match 0.7%; Score 17; DB 4; Length 2355;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 41 GAAATATATCGTGTTC 57
DB 46 GAAATATATCGTGTTC 30
RESULT 85
US-08-318-826A-9
Sequence 9, Application US/08318826A
GENERAL INFORMATION:
APPLICANT: Soreq, Hermona
APPLICANT: Zakut, Haim
APPLICANT: Eckstein, Fritz
TITLE OF INVENTION: Synthetic Antisense
TITLE OF INVENTION: Oligodeoxynucleotides and Pharmaceutical Compositions
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kohn & Associates
STREET: 30500 No. 5891725thwestern Hwy., Suite 410
CITY: Farmington Hills
STATE: Michigan
COUNTRY: US
```

```
ZIP: 48334
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/318,826A
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Kohn, Kenneth I.
REGISTRATION NUMBER: 30,955
REFERENCE/DOCKET NUMBER: 2391.00001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (248) 539-5050
TELEFAX: (248) 539-5055
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2381 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
DEVELOPMENTAL STAGE: fetal
TISSUE TYPE: Brain, Liver
POSITION IN GENOME:
MAP POSITION: 3q26
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 160..1881
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /EC_number=3.1.1.8
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="BCHR"
OTHER INFORMATION: /note="butyrylcholinesterase mature peptide"
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 76..159
FEATURE:
NAME/KEY: mRNA
LOCATION: 1..2381
FEATURE:
NAME/KEY: CDS
LOCATION: 76..1884
US-08-318-826A-9
Query Match 0.7%; Score 17; DB 2; Length 2381;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2300 GATACAAATGATATCAT 2316
DB 1177 GATACAAATGATATCAT 1193
RESULT 86
5215909-13
Patent No. 5215909
APPLICANT: SORBO, HERMONA
TITLE OF INVENTION: HUMAN CHOLINESTERASE GENES
NUMBER OF SEQUENCES: 13
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/572,911
FILING DATE: 15-AUG-1990
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 87,724
FILING DATE: 21-AUG-1987
APPLICATION NUMBER: 875,737
```

FILING DATE: 18-JUN-1986
SEQ ID NO:13
LENGTH: 2400
5215909-13

Query Match 0.7%; Score 17; DB 6; Length 2400;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2300 GATACATAGTATCAT 2316
DB 1177 GATACATAGTATCAT 1193

RESULT 87
US-08-318-826A-8
Sequence 8, Application US/08318826A
Patent No. 5891725
GENERAL INFORMATION:
APPLICANT: Sored, Hermona
APPLICANT: Zakut, Haim
APPLICANT: Eckstein, Fritz
TITLE OF INVENTION: Synthetic Antisense
TITLE OF INVENTION: Oligodeoxynucleotides and Pharmaceutical Compositions
TITLE OF INVENTION: Containing Them
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kohn & Associates
STREET: 30500 No. 5891725thwestern Hwy., Suite 410
CITY: Farmington Hills
STATE: Michigan
COUNTRY: US
ZIP: 48334
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/318.826A
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Kohn, Kenneth I.
REGISTRATION NUMBER: 30,955
REFERENCE/DOCKET NUMBER: 2391.00001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (248) 539-5050
TELEFAX: (248) 539-5055
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 2416 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: Basal ganglion
POSITION IN GENOME:
MAP POSITION: 3q26
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 214..1935
OTHER INFORMATION: /EC number= 3.1.1.8
OTHER INFORMATION: /gene= "BCEH"
OTHER INFORMATION: /note= "butyrylcholinesterase mature peptide"
NAME/KEY: sig_peptide
LOCATION: 130..213
FEATURE:

NAME/KEY: mRNA
LOCATION: 1..2416
FEATURE:
NAME/KEY: CDS
LOCATION: 130..1938
US-08-318-826A-8

Query Match 0.7%; Score 17; DB 2; Length 2416;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2300 GATACATAGTATCAT 2316
DB 1231 GATACATAGTATCAT 1247

RESULT 88
US-09-334-489-1
Sequence 1, Application US/09334489
Patent No. 6291175
GENERAL INFORMATION:
APPLICANT: Pierre Sevigny
APPLICANT: Keith Schepert
APPLICANT: Heiko Wiesbusch
TITLE OF INVENTION: METHODS FOR TREATING A NEUROLOGICAL
TITLE OF INVENTION: DISEASE BY DETERMINING BCEH GENOTYPE
FILE REFERENCE: 08523/013002
CURRENT APPLICATION NUMBER: US/09/334,489
CURRENT FILING DATE: 1999-06-16
PRIOR APPLICATION NUMBER: 60/089,406
PRIOR FILING DATE: 1998-06-18
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 1
LENGTH: 2416
TYPE: DNA
ORGANISM: Homo sapiens
US-09-334-489-1

Query Match 0.7%; Score 17; DB 3; Length 2416;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2300 GATACATAGTATCAT 2316
DB 1231 GATACATAGTATCAT 1247

RESULT 89
US-09-334-489-2
Sequence 2, Application US/09334489
Patent No. 6291175
GENERAL INFORMATION:
APPLICANT: Pierre Sevigny
APPLICANT: Keith Schepert
APPLICANT: Heiko Wiesbusch
TITLE OF INVENTION: METHODS FOR TREATING A NEUROLOGICAL
TITLE OF INVENTION: DISEASE BY DETERMINING BCEH GENOTYPE
FILE REFERENCE: 08523/013002
CURRENT APPLICATION NUMBER: US/09/334,489
CURRENT FILING DATE: 1999-06-16
PRIOR APPLICATION NUMBER: 60/089,406
PRIOR FILING DATE: 1998-06-18
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 2
LENGTH: 2416
TYPE: DNA
ORGANISM: Homo sapiens
US-09-334-489-2

Query Match 0.7%; Score 17; DB 3; Length 2416;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2300 GATACCAATAGTATCAT 2316
|||||
Db 1231 GATACCAATAGTATCAT 1247

RESULT 90
US-08-175-388-3/C
; Sequence 3, Application US/08175388
; Patent No. 5641674
; GENERAL INFORMATION:
; APPLICANT: DEHOUX, PIERRE
; APPLICANT: DAVIES, JULIAN
; TITLE OF INVENTION: Nucleotide Sequences and Proteins
; TITLE OF INVENTION: Conferring Cycloheximide Resistance
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSER: P.C.
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/175,388
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR92/00685
; FILING DATE: 15-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9108906
; FILING DATE: 15-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, No. 5641674man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 660-071-0 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2842 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-175-388-3

Query Match 0.7%; Score 17; DB 1; Length 2842;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1922 AAATTGCTTGTTGA 1938
|||||
Db 1390 AAATTGCTTGTTGA 1374

RESULT 91
US-08-779-620-3/C
; Sequence 3, Application US/08779620
; Patent No. 5830997
; GENERAL INFORMATION:
; APPLICANT: DEHOUX, PIERRE
; APPLICANT: DAVIES, JULIAN
; FILING DATE: 15-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, No. 5641674man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 660-071-0 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2842 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-779-620-3

;; TITLE OF INVENTION: Nucleotide Sequences and Proteins
;; TITLE OF INVENTION: Conferring Cycloheximide Resistance
;; NUMBER OF SEQUENCES: 3
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
;; ADDRESSER: P.C.
;; STREET: 1755 S. Jefferson Davis Highway, Suite 400
;; CITY: Arlington
;; STATE: Virginia
;; COUNTRY: U.S.A.
;; ZIP: 22202
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: IBM PC compatible
;; SOFTWARE: Patent In Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/779,620
;; FILING DATE:
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/175,388
;; FILING DATE: 09-JUN-1994
;; APPLICATION NUMBER: PCT/FR92/00685
;; FILING DATE: 15-JUL-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: FR 9108906
;; FILING DATE: 15-JUL-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: OBLON, No. 5830997man F.
;; REGISTRATION NUMBER: 24,618
;; REFERENCE/DOCKET NUMBER: 660-071-0 PCT
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (703) 413-3000
;; TELEFAX: (703) 413-2220
;; TELEX: 248855 OPAT UR
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2842 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: unknown
;; MOLECULE TYPE: DNA (genomic)
US-08-779-620-3

Query Match 0.7%; Score 17; DB 2; Length 2842;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1922 AAATTGCTTGTTGA 1938
|||||
Db 1390 AAATTGCTTGTTGA 1374

RESULT 92
US-08-818-726-3/C
; Sequence 3, Application US/08818726
; Patent No. 5958687
; GENERAL INFORMATION:
; APPLICANT: DEHOUX, PIERRE
; APPLICANT: DAVIES, JULIAN
; TITLE OF INVENTION: Nucleotide Sequences and Proteins
; TITLE OF INVENTION: Conferring Cycloheximide Resistance
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSER: P.C.
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/918,726
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175,388
FILING DATE: 09-JUN-1994
APPLICATION NUMBER: PCT/FR92/00685
FILING DATE: 15-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 9108906
FILING DATE: 15-JUL-1991
ATTORNEY/AGENT INFORMATION:
NAME: Oblon, No. 5958687man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 660-071-0 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2842 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-818-726-3

Query Match 0.7%; Score 17; DB 2; Length 2842;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1922 AATTGCTGTGTTGA 1938
|||
Db 1390 AATTGCTGTGTTGA 1374

RESULT 93
US-09-613-303-20
Sequence 20, Application US/09613303
Patent No. 6493347
GENERAL INFORMATION:
APPLICANT: Siegel, Marvin
APPLICANT: Chu, N. Randall
APPLICANT: Mizzen, Lee A.
TITLE OF INVENTION: INDUCTION OF A TH1-LIKE RESPONSE IN VITRO
FILE REFERENCE: 12071/002001
CURRENT APPLICATION NUMBER: US/09/613,303
CURRENT FILING DATE: 2000-07-10
PRIOR APPLICATION NUMBER: US 60/143,757
PRIOR FILING DATE: 1999-07-08
NUMBER OF SEQ ID NOS: 55
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 20
LENGTH: 2847
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: fusion sequence
NAME/KEY: CDS
LOCATION: (1)...(2844)
US-09-613-303-20

Query Match 0.7%; Score 17; DB 4; Length 2847;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 422 CAACATCGTGAAGTGG 438

Db 1371 CAACATCGTGAAGTGG 1387
|||

RESULT 94
US-09-484-970B-68/C
Sequence 68, Application US/09484970B
Patent No. 6426186
GENERAL INFORMATION:
APPLICANT: Jones, Karen A.
APPLICANT: Volkmuth, Wayne
APPLICANT: Walker, Michael G.
TITLE OF INVENTION: BONE REMODELING GENES
FILE REFERENCE: PB-0014 US
CURRENT APPLICATION NUMBER: US/09/484,970B
CURRENT FILING DATE: 2000-01-18
NUMBER OF SEQ ID NOS: 172
SOFTWARE: PERL Program
SEQ ID NO 68
LENGTH: 3882
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. 6426186 331276.3CBt
NAME/KEY: unsure
LOCATION: 141
OTHER INFORMATION: a, t, c, g, or other
US-09-484-970B-68

Query Match 0.7%; Score 17; DB 4; Length 3882;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1262 GTTTCACACCTTTAAC 1278
|||
Db 1291 GTTTCACACCTTTAAC 1275

RESULT 95
US-08-955-565A-3
Sequence 3, Application US/08955565A
Patent No. 6331388
GENERAL INFORMATION:
APPLICANT: Malkovsky, Miroslav
APPLICANT: Wells, Andrew
TITLE OF INVENTION: Immune Response Enhancer Therapy
FILE REFERENCE: WARF-02625
CURRENT APPLICATION NUMBER: US/08/955,565A
CURRENT FILING DATE: 1997-10-17
NUMBER OF SEQ ID NOS: 6
SOFTWARE: Patentin version 3.0
SEQ ID NO 3
LENGTH: 4380
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
US-08-955-565A-3

Query Match 0.7%; Score 17; DB 4; Length 4380;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 422 CAACATCGTGAAGTGG 438
|||
Db 1562 CAACATCGTGAAGTGG 1578

RESULT 96
US-09-436-874-1
Sequence 1, Application US/09436874
Patent No. 6521816
GENERAL INFORMATION:
APPLICANT: Froberg, Claus

;; TITLE OF INVENTION: NUCLEIC ACID MOLECULES FROM RICE AND THEIR USE FOR THE
;; FILE REFERENCE: GFB-10
;; CURRENT APPLICATION NUMBER: US/09/436,874
;; EARLIER FILING DATE: 1999-11-09
;; EARLIER FILING DATE: 1998-11-09
;; NUMBER OF SEQ ID NOS: 26
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO: 1
;; LENGTH: 4643
;; TYPE: DNA
;; ORGANISM: Oryza sativa
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (2)..(4375)
;; US-09-436-874-1

Query Match 0.7%; Score 17; DB 4; Length 4643;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 534 AGGATCTGCTGATGCT 550
DB 1623 AGGATCTGCTGATGCT 1639

RESULT 97
US-09-491-356C-2
; Sequence 2, Application US/09491356C
; Patent No. 6566061
; GENERAL INFORMATION:
; APPLICANT: Philibert, Robert A.
; APPLICANT: Gims, Edward I.
; APPLICANT: Delisi, Lynn
; TITLE OF INVENTION: IDENTIFICATION OF POLYMORPHISMS IN THE PCTG4 REGION OF XQ13
; FILE REFERENCE: 9465.6US11
; CURRENT APPLICATION NUMBER: US/09/491,356C
; CURRENT FILING DATE: 2000-01-26
; PRIOR APPLICATION NUMBER: PCT/US99/09365
; PRIOR FILING DATE: 1999-04-29
; PRIOR APPLICATION NUMBER: 60/083,465
; PRIOR FILING DATE: 1998-04-29
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 2
; LENGTH: 6794
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-491-356C-2

Query Match 0.7%; Score 17; DB 4; Length 6794;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 755 AGCAACAAGCAGCAGC 771
DB 6347 AGCAACAAGCAGCAGC 6363

RESULT 98
US-08-276-594A-1/C
; Sequence 1, Application US/08276594A
; Patent No. 5693499
; GENERAL INFORMATION:
; APPLICANT: YONEMURA, Hiroshi
; APPLICANT: TAJIMA, Yoshitaka
; APPLICANT: SUGAWARA, Keishin
; APPLICANT: MASUDA, Kenichi
; TITLE OF INVENTION: PROCESS FOR PREPARING HUMAN COAGULATION
; TITLE OF INVENTION: FACTOR VIII PROTEIN COMPLEX
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Foley & Lardner
;; STREET: 3000 K Street, N.W., Suite 500
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: USA
;; ZIP: 20007-5109
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: IBM PC compatible
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; FILING DATE: 18-JUL-1994
;; APPLICATION NUMBER: US/08/276,594A
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/950,191
;; FILING DATE: 24-SEP-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 243262/1991
;; FILING DATE: 24-SEP-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: WEGNER, Harold C.
;; REGISTRATION NUMBER: 25,258
;; REFERENCE/DOCKET NUMBER: 74129/195/AOPA
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202)672-5300
;; TELEFAX: (202)672-5399
;; TELEX: 904136
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 6999 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1..6996
;; US-08-276-594A-1

Query Match 0.7%; Score 17; DB 1; Length 6999;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1923 AATTGCTTGCTTGAT 1939
DB 3861 AATTGCTTGCTTGAT 3845

RESULT 99
US-08-121-202-1/C
; Sequence 1, Application US/08121202
; Patent No. 5563045
; GENERAL INFORMATION:
; APPLICANT: Pittman, Debra
; APPLICANT: Rehentulla, Alnawaz
; APPLICANT: Wozney, John M.
; APPLICANT: Kaufman, Randal J.
; TITLE OF INVENTION: CHIMERIC PROCOAGULANT PROTEINS
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Legal Affairs, Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/121,202
FILING DATE: 14-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Weinert, M. C.
REGISTRATION NUMBER: 31,544
REFERENCE/DOCKET NUMBER: GI 5195A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1210 X8574
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 7056 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..7053
US-08-121-202-1

Query Match 0.7%; Score 17; DB 1; Length 7056;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1923 AATTGCTTGTTGAT 1939
|||||
Db 3918 AATTGCTTGTTGAT 3902

RESULT 100
5171844-1/c
; Patent No. 5171844
; APPLICANT: VAN COYEN, ALBERT J.J.; PANNKOEK, HANS; VERBEET,
; MARTINUS P.; VAN LEEN, ROBERT W.
; TITLE OF INVENTION: PROTEINS WITH FACTOR VIII ACTIVITY
; PROCESS FOR THEIR PREPARATION USING GENETICALLY-ENGINEERED CELLS
; AND PHARMACEUTICAL COMPOSITIONS CONTAINING THEM
; NUMBER OF SEQUENCES: 12
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/205,226
; FILING DATE: 10-JUN-1988
; SEQ ID NO: 1:
; LENGTH: 8241
5171844-1

Query Match 0.7%; Score 17; DB 6; Length 8241;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1923 AATTGCTTGTTGAT 1939
|||||
Db 3924 AATTGCTTGTTGAT 3908

Search completed: November 14, 2003, 07:51:56
Job time : 179 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 06:26:11 ; Search time 1698 Seconds
(without alignments)
4598.198 Million cell updates/sec

Title: US-09-765-271-55

Perfect score: 2389

Sequence: 1 TTCTTACGAGTTGGAGCTGT.....TAAGTAGGAAATAAATTAAC 2389

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2169961 seqs, 1634102185 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 200 summaries

Database : Published Applications_NA.*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2389	100.0	2389	9	US-09-765-272-55
2	148	6.2	2523	11	US-09-884-465A-3
3	148	6.2	2647	11	US-09-884-465A-4
4	53	2.2	2290	9	US-09-765-272-65
5	53	2.2	2481	11	US-09-769-787-206
6	53	2.2	2639	11	US-09-884-465A-5
7	30	1.3	40	9	US-09-765-272-282
8	27	1.1	37	9	US-09-765-272-281
9	25	1.0	33	11	US-09-884-465A-45
10	25	1.0	34	11	US-09-884-465A-47
11	25	1.0	35	11	US-09-884-465A-174
12	24	1.0	1342	9	US-09-765-272-181
13	24	1.0	1455	11	US-09-769-787-246
14	24	1.0	1455	12	US-09-769-744A-23
15	24	1.0	2528	11	US-09-884-465A-9
16	24	1.0	3120	11	US-09-884-465A-1

17	24	1.0	5048	11	US-09-884-465A-2	Sequence 2, Appli
18	21	0.9	29	11	US-09-884-465A-44	Sequence 44, Appl
19	20	0.8	1146	12	US-09-769-76-23	Sequence 23, Appl
20	20	0.8	2469	12	US-09-769-736-17	Sequence 17, Appl
21	20	0.8	5215	11	US-09-252-088-13	Sequence 13, Appl
22	19	0.8	164	10	US-09-983-965-4749	Sequence 4749, Ap
23	19	0.8	533	12	US-10-027-632-231121	Sequence 231121,
24	19	0.8	533	12	US-10-027-632-231122	Sequence 231122,
25	19	0.8	533	12	US-10-027-632-231123	Sequence 231123,
26	19	0.8	533	12	US-10-027-632-231121	Sequence 231121,
27	19	0.8	533	13	US-10-027-632-231122	Sequence 231122,
28	19	0.8	533	13	US-10-027-632-231123	Sequence 231123,
29	19	0.8	1112	12	US-10-027-632-118539	Sequence 118539,
30	19	0.8	1112	12	US-10-027-632-118540	Sequence 118540,
31	19	0.8	1112	13	US-10-027-632-118539	Sequence 118539,
32	19	0.8	1112	13	US-10-027-632-118540	Sequence 118540,
33	19	0.8	1213	12	US-10-027-632-117222	Sequence 117222,
34	19	0.8	1213	12	US-10-027-632-117223	Sequence 117223,
35	19	0.8	1213	12	US-10-027-632-117224	Sequence 117224,
36	19	0.8	1213	13	US-10-027-632-117222	Sequence 117222,
37	19	0.8	1213	13	US-10-027-632-117223	Sequence 117223,
38	19	0.8	1213	13	US-10-027-632-117224	Sequence 117224,
39	19	0.8	1455	12	US-09-769-736-71	Sequence 71, Appl
40	19	0.8	7493	12	US-10-131-510A-5	Sequence 5, Appli
41	19	0.8	7493	14	US-10-187-319-5	GENERAL INFORMA
42	19	0.8	143306	10	US-09-729-920-3	Sequence 3, Appli
43	18	0.8	324	10	US-09-983-965-5888	Sequence 5888, Ap
44	18	0.8	372	9	US-09-925-302-17	Sequence 17, Appl
45	18	0.8	442	14	US-10-060-036-10	Sequence 10, Appl
46	18	0.8	372	12	US-10-027-632-192700	Sequence 192700,
47	18	0.8	442	13	US-10-027-632-192700	Sequence 192700,
48	18	0.8	461	14	US-09-796-652-4166	Sequence 4166, Ap
49	18	0.8	461	14	US-10-040-862-4166	Sequence 4166, Ap
50	18	0.8	480	12	US-10-027-632-35253	Sequence 35253, A
51	18	0.8	480	12	US-10-027-632-35253	Sequence 35253, A
52	18	0.8	480	12	US-10-027-632-76294	Sequence 76294, A
53	18	0.8	480	12	US-10-027-632-31367	Sequence 31367,
54	18	0.8	480	13	US-10-027-632-35253	Sequence 35253, A
55	18	0.8	480	13	US-10-027-632-35253	Sequence 35253, A
56	18	0.8	480	13	US-10-027-632-76294	Sequence 76294, A
57	18	0.8	480	13	US-10-027-632-31367	Sequence 31367,
58	18	0.8	484	12	US-10-027-632-3008	Sequence 3008, Ap
59	18	0.8	484	13	US-10-027-632-3008	Sequence 3008, Ap
60	18	0.8	536	12	US-10-029-386-11258	Sequence 11258, A
61	18	0.8	536	12	US-10-029-386-11258	Sequence 11258, A
62	18	0.8	563	12	US-10-027-632-284258	Sequence 284258,
63	18	0.8	563	12	US-10-027-632-284258	Sequence 284258,
64	18	0.8	563	13	US-10-027-632-284258	Sequence 284258,
65	18	0.8	577	12	US-10-027-632-81822	Sequence 81822, A
66	18	0.8	577	13	US-10-027-632-81822	Sequence 81822, A
67	18	0.8	608	12	US-10-027-632-185059	Sequence 185059,
68	18	0.8	608	12	US-10-027-632-185059	Sequence 185059,
69	18	0.8	608	13	US-10-027-632-185059	Sequence 185059,
70	18	0.8	674	12	US-10-027-632-27877	Sequence 27877, A
71	18	0.8	674	13	US-10-027-632-27877	Sequence 27877, A
72	18	0.8	924	9	US-09-815-242-9810	Sequence 9810, Ap
73	18	0.8	1519	12	US-09-070-977A-278	Sequence 278, Ap
74	18	0.8	1519	12	US-10-029-386-24974	Sequence 24974, A
75	18	0.8	1429	10	US-10-029-386-25318	Sequence 25318, A
76	18	0.8	1429	12	US-09-833-361-1943	Sequence 1943, Ap
77	18	0.8	1750	9	US-09-925-301-403	Sequence 403, App
78	18	0.8	1803	10	US-09-938-842A-1715	Sequence 1715, Ap
79	18	0.8	2956	9	US-09-764-870-620	Sequence 620, App
80	18	0.8	2956	14	US-10-125-540-620	Sequence 620, App
81	18	0.8	3154	13	US-10-040-090-38	Sequence 38, Appl
82	18	0.8	6705	12	US-10-340-097-5	Sequence 5, Appli
83	18	0.8	6705	12	US-10-336-215-5	Sequence 5, Appli
84	18	0.8	6705	12	US-10-336-219-5	Sequence 2, Appli
85	18	0.8	6619	12	US-10-340-097-2	Sequence 2, Appli
86	18	0.8	6619	12	US-10-336-215-2	Sequence 2, Appli
87	18	0.8	6619	12	US-10-336-219-2	Sequence 2, Appli
88	18	0.8	7488	12	US-10-340-097-1	Sequence 1, Appli
89	18	0.8	7488	12	US-10-336-215-1	Sequence 1, Appli

90	18	0.8	66686	10	US-09-736-960-86
C 91	17	0.7	131	12	US-10-010-942B-18
C 92	17	0.7	146	12	US-10-010-942B-19
C 93	17	0.7	271	10	US-09-878-574-6546
C 94	17	0.7	295	9	US-09-923-876-4422
C 95	17	0.7	316	14	US-10-106-698-2167
C 96	17	0.7	322	9	US-09-765-272-149
C 97	17	0.7	322	14	US-10-060-036-2151
C 98	17	0.7	371	10	US-09-878-574-3595
C 99	17	0.7	375	10	US-09-757-781-26
C 100	17	0.7	402	11	US-09-918-995-36324
C 101	17	0.7	402	12	US-10-010-942B-34
C 102	17	0.7	402	12	US-10-010-942B-35
C 103	17	0.7	419	12	US-10-027-632-83493
C 104	17	0.7	419	13	US-10-027-632-83493
C 105	17	0.7	431	10	US-09-983-965-192
C 106	17	0.7	441	11	US-09-918-995-14286
C 107	17	0.7	458	14	US-10-091-458-22
C 108	17	0.7	461	10	US-09-867-701-4905
C 109	17	0.7	462	11	US-09-918-995-9823
C 110	17	0.7	469	11	US-09-918-995-17823
C 111	17	0.7	470	12	US-10-084-843-216
C 112	17	0.7	470	12	US-10-193-002-211
C 113	17	0.7	475	11	US-09-918-995-3650
C 114	17	0.7	477	12	US-10-027-632-113719
C 115	17	0.7	477	13	US-10-027-632-113719
C 116	17	0.7	479	11	US-09-918-995-5545
C 117	17	0.7	481	10	US-09-867-701-3075
C 118	17	0.7	482	12	US-10-027-632-74403
C 119	17	0.7	482	13	US-10-027-632-74403
C 120	17	0.7	487	11	US-09-918-995-12419
C 121	17	0.7	492	11	US-09-770-961-634
C 122	17	0.7	501	11	US-09-918-995-24817
C 123	17	0.7	512	10	US-09-783-590-8852
C 124	17	0.7	512	12	US-10-027-632-51155
C 125	17	0.7	512	12	US-10-027-632-51156
C 126	17	0.7	512	12	US-10-027-632-58064
C 127	17	0.7	512	12	US-10-027-632-58065
C 128	17	0.7	512	12	US-10-027-632-59098
C 129	17	0.7	512	12	US-10-027-632-59099
C 130	17	0.7	512	13	US-10-027-632-51155
C 131	17	0.7	512	13	US-10-027-632-51156
C 132	17	0.7	512	13	US-10-027-632-58064
C 133	17	0.7	512	13	US-10-027-632-58065
C 134	17	0.7	512	13	US-10-027-632-59098
C 135	17	0.7	512	13	US-10-027-632-59099
C 136	17	0.7	515	12	US-10-028-386-10069
C 137	17	0.7	523	12	US-09-883-227-415
C 138	17	0.7	524	10	US-09-833-381-1769
C 139	17	0.7	543	10	US-09-757-781-27
C 140	17	0.7	544	12	US-10-027-632-4976
C 141	17	0.7	544	13	US-10-027-632-4976
C 142	17	0.7	546	13	US-10-062-254-17
C 143	17	0.7	553	12	US-10-027-632-57333
C 144	17	0.7	553	12	US-10-027-632-294559
C 145	17	0.7	553	13	US-10-027-632-57333
C 146	17	0.7	553	13	US-10-027-632-294559
C 147	17	0.7	566	12	US-10-027-632-184404
C 148	17	0.7	566	13	US-10-027-632-184404
C 149	17	0.7	569	12	US-10-027-632-90772
C 150	17	0.7	569	12	US-10-027-632-90773
C 151	17	0.7	569	12	US-10-027-632-90774
C 152	17	0.7	569	13	US-10-027-632-90772
C 153	17	0.7	569	13	US-10-027-632-90773
C 154	17	0.7	569	13	US-10-027-632-90774
C 155	17	0.7	572	12	US-10-027-632-198206
C 156	17	0.7	572	13	US-10-027-632-198206
C 157	17	0.7	582	12	US-10-027-632-35937
C 158	17	0.7	582	12	US-10-027-632-35937
C 159	17	0.7	582	13	US-10-027-632-35937
C 160	17	0.7	584	12	US-10-027-632-307385
C 161	17	0.7	584	12	US-10-027-632-307385
C 162	17	0.7	584	12	US-10-027-632-307386

Sequence 86, Appl	163	17	0.7	584	12	US-10-027-632-307387
Sequence 18, Appl	164	17	0.7	584	13	US-10-027-632-307385
Sequence 19, Appl	165	17	0.7	584	13	US-10-027-632-307386
Sequence 6546, Ap	166	17	0.7	584	13	US-10-027-632-307387
Sequence 4422, Ap	167	17	0.7	585	12	US-10-027-632-197442
Sequence 2167, Ap	168	17	0.7	585	13	US-10-027-632-197442
Sequence 149, App	169	17	0.7	595	12	US-10-027-632-184901
Sequence 2151, Ap	170	17	0.7	595	13	US-10-027-632-184901
Sequence 3595, Ap	171	17	0.7	599	9	US-09-864-761-6675
Sequence 26, Appl	172	17	0.7	599	12	US-10-027-632-130094
Sequence 36324, A	173	17	0.7	602	13	US-10-027-632-130094
Sequence 34, Appl	174	17	0.7	602	12	US-10-027-632-105942
Sequence 35, Appl	175	17	0.7	602	13	US-10-027-632-105942
Sequence 83493, A	176	17	0.7	604	12	US-10-027-632-126344
Sequence 83493, A	177	17	0.7	604	13	US-10-027-632-126344
Sequence 192, App	178	17	0.7	606	12	US-10-027-632-4107
Sequence 14286, A	179	17	0.7	616	13	US-10-027-632-4107
Sequence 22, Appl	180	17	0.7	614	12	US-10-027-632-5868
Sequence 4905, Ap	181	17	0.7	623	13	US-10-027-632-5868
Sequence 9823, Ap	182	17	0.7	623	12	US-10-113-644-14
Sequence 17823, A	183	17	0.7	624	13	US-10-027-632-187196
Sequence 216, App	184	17	0.7	624	13	US-10-027-632-187196
Sequence 211, App	185	17	0.7	629	12	US-10-027-632-3257
Sequence 313719, Ap	186	17	0.7	629	13	US-10-027-632-3257
Sequence 313719, Ap	187	17	0.7	629	13	US-10-027-632-3257
Sequence 5545, Ap	188	17	0.7	642	12	US-10-027-632-27246
Sequence 3075, Ap	189	17	0.7	642	13	US-10-027-632-27246
Sequence 74403, A	190	17	0.7	650	12	US-10-027-632-198235
Sequence 74403, A	191	17	0.7	650	13	US-10-027-632-198235
Sequence 14191, A	192	17	0.7	662	12	US-10-027-632-132715
Sequence 634, App	193	17	0.7	662	13	US-10-027-632-132715
Sequence 24817, A	194	17	0.7	678	12	US-09-764-886-37
Sequence 8852, Ap	195	17	0.7	686	12	US-10-027-632-263024
Sequence 51155, A	196	17	0.7	686	13	US-10-027-632-263024
Sequence 51156, A	197	17	0.7	694	12	US-10-027-632-110858
Sequence 58065, A	198	17	0.7	694	13	US-10-027-632-110858
Sequence 58065, A	199	17	0.7	720	9	US-09-867-550-767
Sequence 58065, A	200	17	0.7	720	9	US-09-867-550-767

ALIGNMENTS

Sequence 767, App

RESULT 1
US-09-765-272-55
Sequence 55, Application US/09765272
Patent No. US20020061545A1

GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/765,272
FILING DATE: 22-Jan-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512
 INFORMATION FOR SEQ ID NO: 55:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2389 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 SEQUENCE DESCRIPTION: SEQ ID NO: 55:
 us-09-765-272-55

Query Match 100.0%; Score 2389; DB 9; Length 2389;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 TTCTTACGAGTTGGGACTGTATCAAGCTAAGAAACGGTTAAGAAATAATCGTGTCTCTA 60
DB 1 TTCTTACGAGTTGGGACTGTATCAAGCTAAGAAACGGTTAAGAAATAATCGTGTCTCTA 60
QY 61 TATAGTGAAGAAACGAGCGCAAAACGAGAAATTTGACTCTGATGAGGTTAGCA 120
DB 61 TATAGTGAAGAAACGAGCGCAAAACGAGAAATTTGACTCTGATGAGGTTAGCA 120
QY 121 GCGTAAGAGATCAATGCTGAGCAAAATCGTCATCAAGATACAGACCAAGGCTATGTCAC 180
DB 121 GCGTAAGAGATCAATGCTGAGCAAAATCGTCATCAAGATACAGACCAAGGCTATGTCAC 180
QY 181 TTCACATGGCGACCACTATCATTTATTAACAATGTAAGGTTCTTATGACGCTATCATCAG 240
DB 181 TTCACATGGCGACCACTATCATTTATTAACAATGTAAGGTTCTTATGACGCTATCATCAG 240
QY 241 TGAAGAATTACTCATGAAGAATCCAAACTATAGCTAAAGATGAGATATGTTAATGA 300
DB 241 TGAAGAATTACTCATGAAGAATCCAAACTATAGCTAAAGATGAGATATGTTAATGA 300
QY 301 GGTCAAGGTTGATATGTTATCAAGGTAGATGAAATCTATGTTTACCTTAAGGATGC 360
DB 301 GGTCAAGGTTGATATGTTATCAAGGTAGATGAAATCTATGTTTACCTTAAGGATGC 360
QY 361 TGCCCAACGGGATTAACGTCGTAACAAGAGAAATCAATGCAACAAAACAAGACATAG 420
DB 361 TGCCCAACGGGATTAACGTCGTAACAAGAGAAATCAATGCAACAAAACAAGACATAG 420
QY 421 TCAACATGCTGAAGTGAACCTCCAAAGAACATGCTGCTGCTTGGCCTTGGCAGCTTCGA 480
DB 421 TCAACATGCTGAAGTGAACCTCCAAAGAACATGCTGCTGCTTGGCCTTGGCAGCTTCGA 480
QY 481 AGAAGCTATATCAAGATGATGTTATCTTTAATGCTTCTGATATCATAGAGATAC 540
DB 481 AGAAGCTATATCAAGATGATGTTATCTTTAATGCTTCTGATATCATAGAGATAC 540
QY 541 TGGTATGCTTATATCGTTCTCATGAGATCATTAACATTACCTTAAGAAATGAGTT 600
DB 541 TGGTATGCTTATATCGTTCTCATGAGATCATTAACATTACCTTAAGAAATGAGTT 600
QY 601 ATCACTAGCCGAGTTGGCTGTCGAGAAACCTTCATCTGATGTCAGAGAAATCTGTCAA 660
DB 601 ATCACTAGCCGAGTTGGCTGTCGAGAAACCTTCATCTGATGTCAGAGAAATCTGTCAA 660
QY 661 TTCAGAAGACCTATGCGCCGCAAAATAGCGATTAACCTTCAAGAAACAACTGGGTACCTTC 720
DB 661 TTCAGAAGACCTATGCGCCGCAAAATAGCGATTAACCTTCAAGAAACAACTGGGTACCTTC 720
QY 721 TGTAAAGCAATCCAGAACTCAAAATCTTAACAAGCAACAAACAGCAACACTTAACAGTCA 780
DB 721 TGTAAAGCAATCCAGAACTCAAAATCTTAACAAGCAACAAACAGCAACACTTAACAGTCA 780
QY 781 AGCAAGTCAAAGTAATGATGATGATGTTGTTGAAACAGCTCTTACAACTGCTTTGAG 840
DB 781 AGCAAGTCAAAGTAATGATGATGATGTTGTTGAAACAGCTCTTACAACTGCTTTGAG 840

```

```

QY 841 TCAAGCATGTAGATCTGATGCGCTTGTCTTGTATGCACAGCAAAATCACAAGTGCAGC 900
DB 841 TCAAGCATGTAGATCTGATGCGCTTGTCTTGTATGCACAGCAAAATCACAAGTGCAGC 900
QY 901 AGCTAGAGTGTGGAGTGCACACGAGATCATTTACCACTTCACTCTTACTCTCAAT 960
DB 901 AGCTAGAGTGTGGAGTGCACACGAGATCATTTACCACTTCACTCTTACTCTCAAT 960
QY 961 GTCTGAATTTGGAAGAAAGATGCGCTGATTAATCCCTTGGTTATGCTTCAACCAATTC 1020
DB 961 GTCTGAATTTGGAAGAAAGATGCGCTGATTAATCCCTTGGTTATGCTTCAACCAATTC 1020
QY 1021 GGTACCAATTCAGAGCCAGAAACCAAGTCCACACGAGCTCCGGAACCTAGTCCAGG 1080
DB 1021 GGTACCAATTCAGAGCCAGAAACCAAGTCCACACGAGCTCCGGAACCTAGTCCAGG 1080
QY 1081 CCCGCAACCTGACCAATCTTTAAATAGACTCAAAATCTTCTTGTGTTAGTCACTGCT 1140
DB 1081 CCCGCAACCTGACCAATCTTTAAATAGACTCAAAATCTTCTTGTGTTAGTCACTGCT 1140
QY 1141 ACGAAAGTTGGGGAAGGATATGTTACGAAGAAAGGCAATCTCTGTTATGCTTTC 1200
DB 1141 ACGAAAGTTGGGGAAGGATATGTTACGAAGAAAGGCAATCTCTGTTATGCTTTC 1200
QY 1201 GAAAGATTTACCATCTGAACCTGTTAAATCTTGAAGCAAGTTATCAAAACAAGAG 1260
DB 1201 GAAAGATTTACCATCTGAACCTGTTAAATCTTGAAGCAAGTTATCAAAACAAGAG 1260
QY 1261 TGTTCACACACTTAACTGCTAATAAAGAAATGTTGCTCTCTGTCGACCAAGATTTTA 1320
DB 1261 TGTTCACACACTTAACTGCTAATAAAGAAATGTTGCTCTCTGTCGACCAAGATTTTA 1320
QY 1321 TGATTAAGCATATATCTGTTAACTGAGGCTCATTAACCTGTTGTAATTAAGGTCG 1380
DB 1321 TGATTAAGCATATATCTGTTAACTGAGGCTCATTAACCTGTTGTAATTAAGGTCG 1380
QY 1381 TAAATCTGATTTCCAGGCTTGAACAAATTTAAGAGCGTGAATGATGATGACATTA 1440
DB 1381 TAAATCTGATTTCCAGGCTTGAACAAATTTAAGAGCGTGAATGATGATGATGACATTA 1440
QY 1441 TAAAGAAATTTGGTAGATTTATTTGGCAATTCCTAGCACCAATTAACCATCCAGACG 1500
DB 1441 TAAAGAAATTTGGTAGATTTATTTGGCAATTCCTAGCACCAATTAACCATCCAGACG 1500
QY 1501 ACTGGCAACCAATTCCTCAATTTGATATGATGTAAGCAAGAGTTCATTTGCTCAAT 1560
DB 1501 ACTGGCAACCAATTCCTCAATTTGATATGATGTAAGCAAGAGTTCATTTGCTCAAT 1560
QY 1561 AGCTGATTAAGTATCAACGTCAGATGTTTACATTTTGAATGAACATGATATATCAGTGA 1620
DB 1561 AGCTGATTAAGTATCAACGTCAGATGTTTACATTTTGAATGAACATGATATATCAGTGA 1620
QY 1621 TGAAGAGATGCATATGTAACGCTCATATGAGGCCATATGTCATGATTTGGAAGAAATAG 1680
DB 1621 TGAAGAGATGCATATGTAACGCTCATATGAGGCCATATGTCATGATTTGGAAGAAATAG 1680
QY 1681 CCTTCTGATTAAGAAAGTTCAGCTCAAGCTCATCTAAAGAAAGAAAGTTCCTTACC 1740
DB 1681 CCTTCTGATTAAGAAAGTTCAGCTCAAGCTCATCTAAAGAAAGAAAGTTCCTTACC 1740
QY 1741 TCCATCTCCAGACGAGATGTTAAAGCAAAATCCAACTGGAATAGTGCAGCACTATTTA 1800
DB 1741 TCCATCTCCAGACGAGATGTTAAAGCAAAATCCAACTGGAATAGTGCAGCACTATTTA 1800
QY 1801 CAATGTTGTAAGGGGAAAAAGCAATTCACCTGTTGACTTCCATATATGTTAGACA 1860
DB 1801 CAATGTTGTAAGGGGAAAAAGCAATTCACCTGTTGACTTCCATATATGTTAGACA 1860
QY 1861 TACAGTTAAGGTTAAAAACGTTAATTTGATTTCTCATTAAGATCTATTCCATTAATAT 1920
DB 1861 TACAGTTAAGGTTAAAAACGTTAATTTGATTTCTCATTAAGATCTATTCCATTAATAT 1920
QY 1921 TAAATTTGCTTGTGTTGATGATCAACATTAAGCTCAATGCTCAATGCTTATCTTGAAGA 1980

```

Db	1921	TAAATTTGCTGGTTTGATGATCACAATCAAAAGCTCCAAATGGCTATACCTTGGAAAGA	1980
QY	1981	TTTGTGTTGCCGACGATTAAAGTACTACGTAGAAACACCCCTGACGAAAGCTCCACATTTCTAATGA	2040
Db	1981	TTTGTGTTGCCGACGATTAAAGTACTACGTAGAAACACCCCTGACGAAAGCTCCACATTTCTAATGA	2040
QY	2041	TGGATGGGGCAATGCCAGTAGCATGTGTTAGGCAAGAAAGCCACAGTAGAAGATCCAAA	2100
Db	2041	TGGATGGGGCAATGCCAGTAGCATGTGTTAGGCAAGAAAGCCACAGTAGAAGATCCAAA	2100
QY	2101	TAAGAACTTCAAAAGGGAGATGGAAGACCCAGTAGAGAAAACCTCGTGAGCCAGAAAGTCCC	2160
Db	2101	TAAGAACTTCAAAAGGGAGATGGAAGACCCAGTAGAGAAAACCTCGTGAGCCAGAAAGTCCC	2160
QY	2161	TCAAGTAGAGACTGAAAAAGTAGAAGGCCCACTCAAAGAGCAGAAAGTTTTGCTTGGCGAA	2220
Db	2161	TCAAGTAGAGACTGAAAAAGTAGAAGGCCCACTCAAAGAGCAGAAAGTTTTGCTTGGCGAA	2220
QY	2221	AGTAAAGGATTTCTAGTCTGAAAAGCCAAATGCAACAGAAACTAGCTGGTTTACGAAATTA	2280
Db	2221	AGTAAAGGATTTCTAGTCTGAAAAGCCAAATGCAACAGAAACTAGCTGGTTTACGAAATTA	2280
QY	2281	TTTGAAGCTTCAAATTAATGATACAAATAGTATCATGGCAGAGCAGAAAAATTACTTGC	2340
Db	2281	TTTGAAGCTTCAAATTAATGATACAAATAGTATCATGGCAGAGCAGAAAAATTACTTGC	2340
QY	2341	GTGTGTTAAAGGAAGTAATCCCTCATCTGTAGTAAGAAAAAAATTAAC	2389
Db	2341	GTGTGTTAAAGGAAGTAATCCCTCATCTGTAGTAAGAAAAAAATTAAC	2389

```

; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2290 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 65:
US-09-765-272-65
Query Match
Best Local Similarity 100.0%; Score 53; DB 9; Length 2290;
Pred. No. 7,1e-17;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGATTATTCCTTCCTTCGTTATCGTTCAACCAATTGGGTACCAATTCAAG 1035
DB 944 GCTCGATTATTCCTTCCTTCGTTATCGTTCAACCAATTGGGTACCAATTCAAG 996

RESULT 5
US-09-769-787-206
; Sequence 206, Application US/09769787
; Publication No. US20030091577A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Techniques Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PMC/P21129MO
; CURRENT APPLICATION NUMBER: US/09/769,787
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 206
; LENGTH: 2481
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-769-787-206
Query Match
Best Local Similarity 100.0%; Score 53; DB 11; Length 2481;
Pred. No. 7,2e-17;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGATTATTCCTTCCTTCGTTATCGTTCAACCAATTGGGTACCAATTCAAG 1035
DB 1003 GCTCGATTATTCCTTCCTTCGTTATCGTTCAACCAATTGGGTACCAATTCAAG 1055

RESULT 6
US-09-884-465A-5
; Sequence 5, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biotech, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Chailand, Nathalie
; APPLICANT: Quillet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
```

```

; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 2639
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-5
Query Match
Best Local Similarity 100.0%; Score 53; DB 11; Length 2639;
Pred. No. 7,2e-17;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGATTATTCCTTCCTTCGTTATCGTTCAACCAATTGGGTACCAATTCAAG 1035
DB 1113 GCTCGATTATTCCTTCCTTCGTTATCGTTCAACCAATTGGGTACCAATTCAAG 1165

RESULT 7
US-09-765-272-282/c
; Sequence 282, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESS: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 282:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 40 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 282:
US-09-765-272-282
Query Match
Best Local Similarity 100.0%; Score 30; DB 9; Length 40;
Pred. No. 7,2e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2360 CCTCATCTGTAGTAGGAAAAAATAAAC 2389
DB 40 CCTCATCTGTAGTAGGAAAAAATAAAC 11

RESULT 8
US-09-765-272-281
; Sequence 281, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
```

```

; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 281:
;     SEQUENCE CHARACTERISTICS:
;         LENGTH: 37 base pairs
;         TYPE: nucleic acid
;         STRANDEDNESS: double
;         TOPOLOGY: linear
;     SEQUENCE DESCRIPTION: SEQ ID NO: 281:
;
; US-09-765-272-281
;
; Query Match          1.1%; Score 27; DB 9; Length 37;
; Best Local Similarity 100.0%; Pred. No. 0.0028;
; Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY      1  TTCTTACGAGTTGGAGCTGTATCAAGC 27
;         |||||||
; DB      11 TTCTTACGAGTTGGAGCTGTATCAAGC 37
;
; RESULT 9
; US-09-884-465A-45
; Sequence 45, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 45
; LENGTH: 33
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
; US-09-884-465A-45

```

```

; Query Match          1.0%; Score 25; DB 11; Length 33;
; Best Local Similarity 100.0%; Pred. No. 0.032;
; Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY      998 CTTCGTTATCGTTCAACCAATGGG 1022
;         |||||||
; DB      9  CTTCGTTATCGTTCAACCAATGGG 33
;
; RESULT 10
; US-09-884-465A-47
; Sequence 47, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47
; LENGTH: 34
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
; US-09-884-465A-47
;
; Query Match          1.0%; Score 25; DB 11; Length 34;
; Best Local Similarity 100.0%; Pred. No. 0.032;
; Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY      797 GACATGTAGTCTCTTGAACACAGC 821
;         |||||||
; DB      10 GACATGTAGTCTCTTGAACACAGC 34
;
; RESULT 11
; US-09-884-465A-174
; Sequence 174, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 174
; LENGTH: 35
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
; US-09-884-465A-174

```

Query Match 1.0%; Score 25; DB 11; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.032;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 797 GACATGATAGCTCTTGAACAGC 821
Db 11 GACATGATAGCTCTTGAACAGC 35

RESULT 12

US-09-765-272-181
Sequence 181, Application US/09765272
Patent No. US2002061545A1

GENERAL INFORMATION:

APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/765,272
FILING DATE: 22-Jan-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/961,083
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 181:

SEQUENCE CHARACTERISTICS:
LENGTH: 1342 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 181:
US-09-765-272-181

Query Match 1.0%; Score 24; DB 9; Length 1342;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 546 ATGCTATATGCTTCATGAG 569
Db 525 ATGCTATATGCTTCATGAG 548

RESULT 13

US-09-769-787-246

Sequence 246, Application US/09769787
Publication No. US20030091577A1

GENERAL INFORMATION:

APPLICANT: Microbial Technics Limited
APPLICANT: Gilbert, Christophe FG
APPLICANT: Hansbro, Philip M
TITLE OF INVENTION: Proteins
FILE REFERENCE: PWC/P211290
CURRENT APPLICATION NUMBER: US/09/769,787
CURRENT FILING DATE: 2001-01-26

PRIOR APPLICATION NUMBER: GB 9816337.1
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: US 60/125164
PRIOR FILING DATE: 1999-03-19
NUMBER OF SEQ ID NOS: 388
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 246
LENGTH: 1455
TYPE: DNA

ORGANISM: Streptococcus pneumoniae
US-09-769-787-246

Query Match 1.0%; Score 24; DB 11; Length 1455;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 546 ATGCTATATGCTTCATGAG 569
Db 596 ATGCTATATGCTTCATGAG 619

RESULT 14

US-09-769-744A-23

Sequence 23, Application US/09769744A
Publication No. US20030134407A1

GENERAL INFORMATION:

APPLICANT: Le Page, Richard WF
APPLICANT: Wells, Jeremy M
APPLICANT: Hamifly, Sean B
APPLICANT: Hansbro, Philip M
TITLE OF INVENTION: Proteins
FILE REFERENCE: PWC/P211220
CURRENT APPLICATION NUMBER: US/09/769,744A
CURRENT FILING DATE: 2001-01-26

PRIOR APPLICATION NUMBER: PCT/GB99/02452
PRIOR FILING DATE: 1999-07-27
PRIOR APPLICATION NUMBER: GB 9816336.3
PRIOR FILING DATE: 1998-07-27

PRIOR FILING DATE: 1999-03-19
NUMBER OF SEQ ID NOS: 196
SOFTWARE: Patent In Ver. 2.1

LENGTH: 1455
TYPE: DNA
ORGANISM: Streptococcus pneumoniae
US-09-769-744A-23

Query Match 1.0%; Score 24; DB 12; Length 1455;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 546 ATGCTATATGCTTCATGAG 569
Db 596 ATGCTATATGCTTCATGAG 619

RESULT 15

US-09-884-465A-9

Sequence 9, Application US/09884465A
Publication No. US2003007293A1

GENERAL INFORMATION:

APPLICANT: Shire Biochem, Inc.
APPLICANT: Hamel, Josee
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charlet, Nathalie
TITLE OF INVENTION: Streptococcus Antigenes
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683

PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 9
LENGTH: 2528
TYPE: DNA
ORGANISM: Streptococcus pneumoniae
US-09-884-465A-9

Query Match 1.0%; Score 24; DB 11; Length 2528;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 546 ATGCTTATATCGTTCCTCATGAG 569
DB 536 ATGCTTATATCGTTCCTCATGAG 559

RESULT 16
US-09-884-465A-1
Sequence 1, Application US/09884465A
Publication No. US20030077293A1
GENERAL INFORMATION:
APPLICANT: Shire Biochem, Inc.
APPLICANT: Hamel, Josee
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 3120
TYPE: DNA
ORGANISM: Streptococcus pneumoniae
US-09-884-465A-1

Query Match 1.0%; Score 24; DB 11; Length 3120;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 546 ATGCTTATATCGTTCCTCATGAG 569
DB 596 ATGCTTATATCGTTCCTCATGAG 619

RESULT 17
US-09-884-465A-2
Sequence 2, Application US/09884465A
Publication No. US20030077293A1
GENERAL INFORMATION:
APPLICANT: Shire Biochem, Inc.
APPLICANT: Hamel, Josee
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 5048

TYPE: DNA
ORGANISM: Streptococcus pneumoniae
US-09-884-465A-2

Query Match 1.0%; Score 24; DB 11; Length 5048;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 546 ATGCTTATATCGTTCCTCATGAG 569
DB 2372 ATGCTTATATCGTTCCTCATGAG 2395

RESULT 18
US-09-884-465A-44
Sequence 44, Application US/09884465A
Publication No. US20030077293A1
GENERAL INFORMATION:
APPLICANT: Shire Biochem, Inc.
APPLICANT: Hamel, Josee
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 44
LENGTH: 29
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism
US-09-884-465A-44

Query Match 0.9%; Score 21; DB 11; Length 29;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 621 CTGCAGAGCCTTCCTATCTG 641
DB 9 CTGCAGAGCCTTCCTATCTG 29

RESULT 19
US-09-769-736-23
Sequence 23, Application US/09769736
Publication No. US20030138775A1
GENERAL INFORMATION:
APPLICANT: Microbial Technics Limited
APPLICANT: Le Page, Richard WF
APPLICANT: Wells, Jeremy M
APPLICANT: Hanniffy, Sean B
TITLE OF INVENTION: Proteins
FILE REFERENCE: PWC/P21089wo
CURRENT APPLICATION NUMBER: US/09/769,736
CURRENT FILING DATE: 2003-02-14
PRIOR APPLICATION NUMBER: GB 9816335.5
PRIOR FILING DATE: 1998-07-27
PRIOR APPLICATION NUMBER: US 60/125163
PRIOR FILING DATE: 1999-03-19
NUMBER OF SEQ ID NOS: 212
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 23
LENGTH: 1146
TYPE: DNA
ORGANISM: Streptococcus agalactiae
US-09-769-736-23

Query Match 0.8%; Score 20; DB 12; Length 1146;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 479 CAAGACGCTATCTACTACAGA 498
|||||
DB 571 CAAGACGCTATCTACTACAGA 590

RESULT 20

US-09-769-736-17
; Sequence 17, Application US/09769736
; Publication No. US2003013875A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Le Page, Richard WF
; APPLICANT: Wells, Jeremy M
; APPLICANT: Hamilfy, Sean B
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21089wo
; CURRENT APPLICATION NUMBER: US/09/769,736
; PRIOR FILING DATE: 2003-02-14
; PRIOR APPLICATION NUMBER: GB 9816335.5
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125163
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 17
; LENGTH: 2469
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae
US-09-769-736-17

Query Match 0.8%; Score 20; DB 12; Length 2469;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 479 CAAGACGCTATCTACTACAGA 498
|||||
DB 571 CAAGACGCTATCTACTACAGA 590

RESULT 21

US-09-252-088-13
; Sequence 13, Application US/09252088
; Publication No. US20030031682A1
; GENERAL INFORMATION:
; APPLICANT: BRODEUR, Bernard R.
; APPLICANT: RIOUX, Clment
; APPLICANT: BOYER, Martine
; APPLICANT: CHARLEBOIS, Isabelle
; APPLICANT: HAMEL, Jose
; APPLICANT: MARTIN, Denis
; TITLE OF INVENTION: NOVEL GROUP B STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 8331-9002
; CURRENT APPLICATION NUMBER: US/09/252,088
; PRIOR FILING DATE: 1999-02-18
; EARLIER APPLICATION NUMBER: US/60/075,425
; PRIOR FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 13
; LENGTH: 5215
; TYPE: DNA
; ORGANISM: group B streptococcus

FEATURE:
NAME/KEY: CDS
LOCATION: (3)..(122)
FEATURE:
NAME/KEY: CDS
LOCATION: (133)..(2511) *

FEATURE:
NAME/KEY: CDS
LOCATION: (367)..(2511)

NAME/KEY: CDS
LOCATION: Complement ((2716)..(2946))
FEATURE:
NAME/KEY: CDS
LOCATION: Complement ((2995)..(3252))
FEATURE:
NAME/KEY: CDS
LOCATION: Complement ((3299)..(3676))
FEATURE:
NAME/KEY: CDS
LOCATION: Complement ((3837)..(4124))
FEATURE:
NAME/KEY: CDS
LOCATION: Complement ((4351)..(5214))
US-09-252-088-13

Query Match 0.8%; Score 20; DB 11; Length 5215;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 479 CAAGACGCTATCTACTACAGA 498
|||||
DB 616 CAAGACGCTATCTACTACAGA 635

RESULT 22

US-09-983-965-4749/c
; Sequence 4749, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 37-21 (10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; PRIOR FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1998-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 4749
; LENGTH: 164
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 17-LIB34-068-Q1-E1-E1
US-09-983-965-4749

Query Match 0.8%; Score 19; DB 10; Length 164;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1112 TCAATTCCTTTGTTA 1130
|||||
DB 83 TCAATTCCTTTGTTA 65

RESULT 23

US-10-027-632-231121
; Sequence 231121, Application US/10027632
; Publication No. US20030204075A8
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome

```
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 231121
LENGTH: 533
TYPE: DNA
ORGANISM: Human
US-10-027-632-231121

Query Match      0.8%; Score 19; DB 12; Length 533;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1212 CATCTGAAACTGTTAAAAA 1230
Db      31 CATCTGAAACTGTTAAAAA 49

RESULT 24
US-10-027-632-231122
Sequence 231122, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 231122
LENGTH: 533
TYPE: DNA
ORGANISM: Human
US-10-027-632-231122

Query Match      0.8%; Score 19; DB 12; Length 533;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1212 CATCTGAAACTGTTAAAAA 1230
Db      31 CATCTGAAACTGTTAAAAA 49
```

```
Db      31 CATCTGAAACTGTTAAAAA 49

RESULT 25
US-10-027-632-231123
Sequence 231123, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 231121
LENGTH: 533

QY      1212 CATCTGAAACTGTTAAAAA 1230
Db      31 CATCTGAAACTGTTAAAAA 49

RESULT 26
US-10-027-632-231121
Sequence 231121, Application US/10027632
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 231121
LENGTH: 533

Query Match      0.8%; Score 19; DB 12; Length 533;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```



```
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-231121

Query Match
Best Local Similarity 100.0%; Score 19; DB 13; Length 533;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1212 CATCTGAACCTGTTAAAA 1230
Db      31 CATCTGAACCTGTTAAAA 49

RESULT 27
US-10-027-632-231122
; Sequence 231122, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 231122
; LENGTH: 533
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-231122

Query Match
Best Local Similarity 100.0%; Score 19; DB 13; Length 533;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1212 CATCTGAACCTGTTAAAA 1230
Db      31 CATCTGAACCTGTTAAAA 49

RESULT 28
US-10-027-632-231123
; Sequence 231123, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
```

```
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 231123
; LENGTH: 533
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-231123

Query Match
Best Local Similarity 100.0%; Score 19; DB 13; Length 533;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1212 CATCTGAACCTGTTAAAA 1230
Db      31 CATCTGAACCTGTTAAAA 49

RESULT 29
US-10-027-632-118539
; Sequence 118539, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 118539
; LENGTH: 1112
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-118539

Query Match
Best Local Similarity 100.0%; Score 19; DB 12; Length 1112;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1513 AAATTCGAATTCAGTAT 1531
Db      891 AAATTCGAATTCAGTAT 909

RESULT 30
US-10-027-632-118540
; Sequence 118540, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
```

```

; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 118540
; LENGTH: 1112
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-118540
```

```

Query Match          0.8%; Score 19; DB 12; Length 1112;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1513 AAATTCGAATTGAGTAT 1531
      |||||
Db      891 AAATTCGAATTGAGTAT 909
```

```

RESULT 31
US-10-027-632-118539
; Sequence 118539, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827,129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 118539
; LENGTH: 1112
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-118539
```

```

Query Match          0.8%; Score 19; DB 13; Length 1112;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1513 AAATTCGAATTGAGTAT 1531
      |||||
Db      891 AAATTCGAATTGAGTAT 909
```

```

RESULT 32
US-10-027-632-118540
; Sequence 118540, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827,129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 118540
; LENGTH: 1112
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-118540
```

```

Query Match          0.8%; Score 19; DB 13; Length 1112;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1513 AAATTCGAATTGAGTAT 1531
      |||||
Db      891 AAATTCGAATTGAGTAT 909
```

```

RESULT 33
US-10-027-632-117222
; Sequence 117222, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827,129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 117222
; LENGTH: 1213
; TYPE: DNA
; ORGANISM: Human
```

US-10-027-632-117222

Query Match 0.8%; Score 19; DB 12; Length 1213;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1759 TGTAAAGCAATCCAACT 1777

Db 776 TGTAAAGCAATCCAACT 794

RESULT 34

US-10-027-632-117223
; Sequence 117223, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:

; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 117223
; LENGTH: 1213
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-117223

Query Match 0.8%; Score 19; DB 12; Length 1213;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1759 TGTAAAGCAATCCAACT 1777

Db 776 TGTAAAGCAATCCAACT 794

RESULT 35

US-10-027-632-117224
; Sequence 117224, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:

; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 117224
; LENGTH: 1213
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-117224

QY 1759 TGTAAAGCAATCCAACT 1777

Db 776 TGTAAAGCAATCCAACT 794

Query Match 0.8%; Score 19; DB 12; Length 1213;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1759 TGTAAAGCAATCCAACT 1777

Db 776 TGTAAAGCAATCCAACT 794

RESULT 36

US-10-027-632-117222
; Sequence 117222, Application US/10027632
; GENERAL INFORMATION:

; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 117222
; LENGTH: 1213
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-117222

Query Match 0.8%; Score 19; DB 13; Length 1213;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1759 TGTAAAGCAATCCAACT 1777

Db 776 TGTAAAGCAATCCAACT 794

RESULT 37

US-10-027-632-117223
; Sequence 117223, Application US/10027632
; GENERAL INFORMATION:

; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30

```

; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 117223
; LENGTH: 1213
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-117223

```

```

Query Match          0.8%; Score 19; DB 13; Length 1213;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1759 TGTAAAGCAATCCACT 1777
      |||||
Db      776 TGTAAAGCAATCCACT 794

```

```

RESULT 38
US-10-027-632-117224
; Sequence 117224, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827,129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 117224
; LENGTH: 1213
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-117224

```

```

Query Match          0.8%; Score 19; DB 13; Length 1213;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1759 TGTAAAGCAATCCACT 1777
      |||||
Db      776 TGTAAAGCAATCCACT 794

```

RESULT 39

```

US-09-769-736-71
; Sequence 71, Application US/09769736
; Publication No. US2003013875A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Le Page, Richard WF
; APPLICANT: Wells, Jeremy M
; APPLICANT: Hamniffy, Sean B
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PMC/P21089WC
; CURRENT APPLICATION NUMBER: US/09/769,736
; CURRENT FILING DATE: 2003-02-14
; PRIOR APPLICATION NUMBER: GB 9816335.5
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125163
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 71
; LENGTH: 1455
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae
US-09-769-736-71

```

```

Query Match          0.8%; Score 19; DB 12; Length 1455;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1583 GATGTTACATTTTGATG 1601
      |||||
Db      343 GATGTTACATTTTGATG 361

```

```

RESULT 40
US-10-131-510A-5
; Sequence 5, Application US/10131510A
; Publication No. US20030166536A1
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S
; TITLE OF INVENTION: Modified Factor VIII
; FILE REFERENCE: 75-957
; CURRENT APPLICATION NUMBER: US/10/131,510A
; CURRENT FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: U.S. 09/315,179
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: U.S. 09/037,601
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: U.S. 08/670,707
; PRIOR FILING DATE: 1996-06-26
; PRIOR APPLICATION NUMBER: PCT/US97/11155
; PRIOR FILING DATE: 1997-06-26
; PRIOR APPLICATION NUMBER: PCT/US94/13200
; PRIOR FILING DATE: 1994-11-15
; PRIOR APPLICATION NUMBER: U.S. 08/212,133
; PRIOR FILING DATE: 1994-03-11
; PRIOR APPLICATION NUMBER: U.S. 07/864,004
; PRIOR FILING DATE: 1992-04-07
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 5
; LENGTH: 7493
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-131-510A-5

```

```

Query Match          0.8%; Score 19; DB 12; Length 7493;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      2347 AAAAGAGTAATCCTCA 2365
      |||||
Db      206 AAAAGAGTAATCCTCA 224

```

RESULT 41
US-10-187-319-5
GENERAL INFORMATION:
APPLICANT: Lollar, John S.
TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
STREET: 5370 Manhattan Circle Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: USA
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/187,319
FILING DATE: 27-Aug-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/523,656
FILING DATE: 2000-03-10
APPLICATION NUMBER: US 09/037,601
FILING DATE: 1998-03-10
APPLICATION NUMBER: WO PCT/US97/11155
FILING DATE: 1997-06-26
APPLICATION NUMBER: US 08/670,707
FILING DATE: 1996-06-26
ATTORNEY/AGENT INFORMATION:
NAME: Greenlee, Lorraine L.
REGISTRATION NUMBER: 27,894
REFERENCE/DOCKET NUMBER: 75-95K
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/499-8080
TELEFAX: 303/499-8089
TITLE: Sequence of the murine Factor VIII cDNA
JOURNAL: Genomics
VOLUME: 16
PAGES: 374-379
DATE: 1993
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-10-187-319-5
Query Match 0.8%; Score 19; DB 14; Length 7493;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2347 AAAAGAGTAATCCTTCA 2355
DB 206 AAAAGAGTAATCCTTCA 224
RESULT 42
US-09-729-920-3/c
Sequence 3, Application US/09799920
Patent No. US20020103115A1
GENERAL INFORMATION:
APPLICANT: GUEGLER, Karl et al
TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
FILE REFERENCE: CL000858
CURRENT APPLICATION NUMBER: US/09/729,920
CURRENT FILING DATE: 2000-12-06
NUMBER OF SEQ ID NOS: 5
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 143306
TYPE: DNA

ORGANISM: Human
US-09-729-920-3
Query Match 0.8%; Score 19; DB 10; Length 143306;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 321 TCAGGTAGATGGAATA 339
DB 18852 TCAGGTAGATGGAATA 18834
RESULT 43
US-09-983-965-5888/c
Sequence 5888, Application US/09983965
Patent No. US20020137160A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 37-21(10297)C
CURRENT APPLICATION NUMBER: US/09/983,965
CURRENT FILING DATE: 2001-10-26
PRIOR APPLICATION NUMBER: US 09/465,231
PRIOR FILING DATE: 1999-12-15
PRIOR APPLICATION NUMBER: US 60/113,678
PRIOR FILING DATE: 1998-12-17
NUMBER OF SEQ ID NOS: 5912
SEQ ID NO 5888
LENGTH: 324
TYPE: DNA
ORGANISM: Bos taurus
FEATURE:
OTHER INFORMATION: Clone ID: 63-LIB34-039-Q1-E1-H4
US-09-983-965-5888
Query Match 0.8%; Score 18; DB 10; Length 324;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1101 TTAATAGACTCAATT 1118
DB 108 TTAATAGACTCAATT 91
RESULT 44
US-09-925-302-17
Sequence 17, Application US/09925302
Patent No. US2002004941A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA104
CURRENT APPLICATION NUMBER: US/09/925,302
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05918
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1998-03-12
NUMBER OF SEQ ID NOS: 896
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 17
LENGTH: 372
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (357)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature

LOCATION: (367)
OTHER INFORMATION: n equals a,t,g, or c
US-09-925-302-17

Query Match 0.8%; Score 18; DB 9; Length 372;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2195 AAGAGCAGAGCTTTG 2212
|||||
DB 99 AAGAGCAGAGCTTTG 116

RESULT 45
US-10-060-036-10
Sequence 10, Application US/10060036
Publication No. US20030073144A1
GENERAL INFORMATION:
APPLICANT: Benson, Darin R.
APPLICANT: Kalos, Michael D.
APPLICANT: Lodes, Michael J.
APPLICANT: Persing, David H.
APPLICANT: Hepier, William T.
APPLICANT: Jiang, Yugu

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
FILE REFERENCE: 210121.566

CURRENT APPLICATION NUMBER: US/10/060,036
CURRENT FILING DATE: 2002-01-30
NUMBER OF SEQ ID NOS: 4560

SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10

LENGTH: 372
TYPE: DNA

ORGANISM: Homo sapiens
US-10-060-036-10

Query Match 0.8%; Score 18; DB 14; Length 372;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 232 TATCATCAGTGAAGATT 249
|||||
DB 228 TATCATCAGTGAAGATT 245

RESULT 46
US-10-027-632-192700
Sequence 192700, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129

CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 192700
LENGTH: 442
TYPE: DNA
ORGANISM: Human
US-10-027-632-192700

Query Match 0.8%; Score 18; DB 12; Length 442;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 754 AAGCAACAGCAACAC 771
|||||
DB 237 AAGCAACAGCAACAC 254

RESULT 47
US-10-027-632-192700
Sequence 192700, Application US/10027632
Publication No. US20030073144A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129

CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 192700
LENGTH: 442
TYPE: DNA

ORGANISM: Human
US-10-027-632-192700

Query Match 0.8%; Score 18; DB 13; Length 442;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 754 AAGCAACAGCAACAC 771
|||||
DB 237 AAGCAACAGCAACAC 254

RESULT 48
US-09-796-692-4166/C
Sequence 4166, Application US/09796692
Publication No. US20020198362A1
GENERAL INFORMATION:
APPLICANT: Gaiger, Alexander
APPLICANT: Algate, Paul A.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
FILE REFERENCE: 2077.001200

CURRENT APPLICATION NUMBER: US/09/796,692
CURRENT FILING DATE: 2001-03-01
PRIOR APPLICATION NUMBER: 60/186,126

PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 60/190,479
PRIOR FILING DATE: 2000-03-17

```
;; PRIOR APPLICATION NUMBER: 60/200,545
;; PRIOR FILING DATE: 2000-04-27
;; PRIOR APPLICATION NUMBER: 60/200,303
;; PRIOR FILING DATE: 2000-04-28
;; PRIOR APPLICATION NUMBER: 60/200,779
;; PRIOR FILING DATE: 2000-04-28
;; PRIOR APPLICATION NUMBER: 60/200,999
;; PRIOR FILING DATE: 2000-05-01
;; PRIOR APPLICATION NUMBER: 60/202,084
;; PRIOR FILING DATE: 2000-05-04
;; PRIOR APPLICATION NUMBER: 60/206,201
;; PRIOR FILING DATE: 2000-05-22
;; PRIOR APPLICATION NUMBER: 60/218,950
;; PRIOR FILING DATE: 2000-07-14
;; PRIOR APPLICATION NUMBER: 60/222,903
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: 60/223,416
;; PRIOR FILING DATE: 2000-08-04
;; PRIOR APPLICATION NUMBER: 60/223,378
;; PRIOR FILING DATE: 2000-08-07
;; NUMBER OF SEQ ID NOS: 9597
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO: 4166
;; LENGTH: 461
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-09-765-692-4166
```

```
Query Match      0.8%; Score 18; DB 10; Length 461;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1740 CTCGATCTCCAGCGCAG 1757
          |||||
Db      122 CTCGATCTCCAGCGCAG 105
```

```
RESULT 49
US-10-040-862-4166/c
;; Sequence 4166, Application US/10040862
;; Publication No. US20030078396A1
;; GENERAL INFORMATION:
;; APPLICANT: Galiger, Alexander
;; APPLICANT: Algate, Paul A.
;; APPLICANT: Mannion, Jane
;; APPLICANT: Retter, Marc
;; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
;; FILE REFERENCE: 014058-013520US
;; CURRENT APPLICATION NUMBER: US/10/040,862
;; CURRENT FILING DATE: 2001-11-06
;; PRIOR APPLICATION NUMBER: US 60/186,126
;; PRIOR FILING DATE: 2000-03-01
;; PRIOR APPLICATION NUMBER: US 60/190,479
;; PRIOR FILING DATE: 2000-03-17
;; PRIOR APPLICATION NUMBER: US 60/200,545
;; PRIOR FILING DATE: 2000-04-27
;; PRIOR APPLICATION NUMBER: US 60/200,303
;; PRIOR FILING DATE: 2000-04-28
;; PRIOR APPLICATION NUMBER: US 60/200,779
;; PRIOR FILING DATE: 2000-04-28
;; PRIOR APPLICATION NUMBER: US 60/200,999
;; PRIOR FILING DATE: 2000-05-01
;; PRIOR APPLICATION NUMBER: US 60/202,084
;; PRIOR FILING DATE: 2000-05-04
;; PRIOR APPLICATION NUMBER: US 60/206,201
;; PRIOR FILING DATE: 2000-05-22
;; PRIOR APPLICATION NUMBER: US 60/218,950
;; PRIOR FILING DATE: 2000-07-14
;; PRIOR APPLICATION NUMBER: US 60/222,903
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: US 60/223,416
```

```
;; PRIOR FILING DATE: 2000-08-04
;; PRIOR APPLICATION NUMBER: US 60/223,378
;; PRIOR FILING DATE: 2000-08-07
;; PRIOR APPLICATION NUMBER: US 09/796,692
;; PRIOR FILING DATE: 2001-03-01
;; NUMBER OF SEQ ID NOS: 10467
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO: 4166
;; LENGTH: 461
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-10-040-862-4166
```

```
Query Match      0.8%; Score 18; DB 14; Length 461;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1740 CTCGATCTCCAGCGCAG 1757
          |||||
Db      122 CTCGATCTCCAGCGCAG 105
```

```
RESULT 50
US-10-027-632-35253
;; Sequence 35253, Application US/10027632
;; Publication No. US20030204075A9
;; GENERAL INFORMATION:
;; APPLICANT: Wang, David G.
;; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;; FILE REFERENCE: 108827.129
;; CURRENT APPLICATION NUMBER: US/10/027,632
;; CURRENT FILING DATE: 2002-04-30
;; PRIOR APPLICATION NUMBER: US 60/218,006
;; PRIOR FILING DATE: 2000-07-12
;; PRIOR APPLICATION NUMBER: US 60/198,676
;; PRIOR FILING DATE: 2000-04-20
;; PRIOR APPLICATION NUMBER: US 60/193,483
;; PRIOR FILING DATE: 2000-03-29
;; PRIOR APPLICATION NUMBER: US 60/185,218
;; PRIOR FILING DATE: 2000-02-24
;; PRIOR APPLICATION NUMBER: US 60/167,363
;; PRIOR FILING DATE: 1999-11-23
;; PRIOR APPLICATION NUMBER: US 60/156,358
;; PRIOR FILING DATE: 1999-09-28
;; PRIOR APPLICATION NUMBER: US 60/146,002
;; PRIOR FILING DATE: 1999-08-09
;; NUMBER OF SEQ ID NOS: 325720
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO: 35253
;; LENGTH: 480
;; TYPE: DNA
;; ORGANISM: Human
US-10-027-632-35253

Query Match      0.8%; Score 18; DB 12; Length 480;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      702 GACCAACTGGGTACTT 719
          |||||
Db      248 GACCAACTGGGTACTT 265

RESULT 51
US-10-027-632-75653
;; Sequence 75653, Application US/10027632
;; Publication No. US20030204075A9
;; GENERAL INFORMATION:
;; APPLICANT: Wang, David G.
;; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;; FILE REFERENCE: 108827.129
```

```

; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 75653
; LENGTH: 480
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-75653

Query Match          0.8%; Score 18; DB 12; Length 480;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      702 GAACAACTGGGTACCTT 719
      |||||
Db      248 GAACAACTGGGTACCTT 265

RESULT 52
US-10-027-632-76294
; Sequence 76294, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 76294
; LENGTH: 480
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-76294

Query Match          0.8%; Score 18; DB 12; Length 480;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      702 GAACAACTGGGTACCTT 719
      |||||
Db      248 GAACAACTGGGTACCTT 265
```

```

RESULT 53
US-10-027-632-313657
; Sequence 313657, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 313657
; LENGTH: 480
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-313657

Query Match          0.8%; Score 18; DB 12; Length 480;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      702 GAACAACTGGGTACCTT 719
      |||||
Db      248 GAACAACTGGGTACCTT 265

RESULT 54
US-10-027-632-35253
; Sequence 35253, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35253
; LENGTH: 480
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-35253
```



```
; ORGANISM: Human
US-10-027-632-35253

Query Match
Best Local Similarity 0.8%; Score 18; DB 13; Length 480;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 702 GAACAACCTGGGTACCTT 719
   |||||
   |||||
   |||||
   |||||
   |||||
Db 248 GAACAACCTGGGTACCTT 265

RESULT 55
US-10-027-632-75653
; Sequence 75653, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 75653
; LENGTH: 480
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-75653

Query Match
Best Local Similarity 0.8%; Score 18; DB 13; Length 480;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 702 GAACAACCTGGGTACCTT 719
   |||||
   |||||
   |||||
   |||||
   |||||
Db 248 GAACAACCTGGGTACCTT 265

RESULT 56
US-10-027-632-76294
; Sequence 76294, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
```

```
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 76294
; LENGTH: 480
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-76294

Query Match
Best Local Similarity 0.8%; Score 18; DB 13; Length 480;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 702 GAACAACCTGGGTACCTT 719
   |||||
   |||||
   |||||
   |||||
   |||||
Db 248 GAACAACCTGGGTACCTT 265

RESULT 57
US-10-027-632-313657
; Sequence 313657, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 313657
; LENGTH: 480
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-313657

Query Match
Best Local Similarity 0.8%; Score 18; DB 13; Length 480;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 702 GAACAACCTGGGTACCTT 719
   |||||
   |||||
   |||||
   |||||
   |||||
Db 248 GAACAACCTGGGTACCTT 265

RESULT 58
US-10-027-632-3008
; Sequence 3008, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
```

PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3008
LENGTH: 484
TYPE: DNA
ORGANISM: Human
US-10-027-632-3008

Query Match 0.8%; Score 18; DB 12; Length 484;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 702 GACCAACTGGGTACTT 719
|||||
Db 244 GACCAACTGGGTACTT 261

RESULT 59
US-10-027-632-3008
Sequence 3008, Application US/10027632
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3008
LENGTH: 484
TYPE: DNA
ORGANISM: Human
US-10-027-632-3008

Query Match 0.8%; Score 18; DB 13; Length 484;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 702 GACCAACTGGGTACTT 719
|||||
Db 244 GACCAACTGGGTACTT 261

RESULT 60

US-10-029-386-11258
Sequence 11258, Application US/10029386
Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
FILE REFERENCE: AECOMICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 11258
LENGTH: 536
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO D87009.1
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.38
OTHER INFORMATION: NT HIT: AF223676.1, EVALUATE 6.00e-10
OTHER INFORMATION: SWISSPROT HIT: Q9HNE4, EVALUATE 2.20e+00
OTHER INFORMATION: EST_HUMAN HIT: BG473533.1, EVALUATE 1.00e-125
US-10-029-386-11258

Query Match 0.8%; Score 18; DB 12; Length 536;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1111 CTCAAATCTCTTTGGT 1128
|||||
Db 218 CTCAAATCTCTTTGGT 235

RESULT 61
US-10-027-632-284258/c
Sequence 284258, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 284258
LENGTH: 563
TYPE: DNA
ORGANISM: Human
US-10-027-632-284258

Query Match 0.8%; Score 18; DB 12; Length 563;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 275 CTAAAGATGAGATATT 292

```
Db          482 CTTAAAGATGAGGATATT 465
|||||
RESULT 62
US-10-027-632-284259/c
; Sequence 284259, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 284259
; TYPE: DNA
; LENGTH: 563
; ORGANISM: Human
US-10-027-632-284259

Query Match          0.8%; Score 18; DB 12; Length 563;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY          275 CTTAAAGATGAGGATATT 292
|||||
Db          482 CTTAAAGATGAGGATATT 465
|||||

RESULT 63
US-10-027-632-284258/c
; Sequence 284258, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 284258
```

```
; LENGTH: 563
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-284258

Query Match          0.8%; Score 18; DB 13; Length 563;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY          275 CTTAAAGATGAGGATATT 292
|||||
Db          482 CTTAAAGATGAGGATATT 465
|||||

RESULT 64
US-10-027-632-284259/c
; Sequence 284259, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 284259
; TYPE: DNA
; LENGTH: 563
; ORGANISM: Human
US-10-027-632-284259

Query Match          0.8%; Score 18; DB 13; Length 563;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY          275 CTTAAAGATGAGGATATT 292
|||||
Db          482 CTTAAAGATGAGGATATT 465
|||||

RESULT 65
US-10-027-632-81822
; Sequence 81822, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
```

```
;; PRIOR FILING DATE: 2000-02-24
;; PRIOR APPLICATION NUMBER: US 60/167,363
;; PRIOR FILING DATE: 1999-11-23
;; PRIOR APPLICATION NUMBER: US 60/156,358
;; PRIOR FILING DATE: 1999-09-28
;; PRIOR APPLICATION NUMBER: US 60/146,002
;; PRIOR FILING DATE: 1999-08-09
;; NUMBER OF SEQ ID NOS: 325720
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO: 81822
;; LENGTH: 577
;; TYPE: DNA
;; ORGANISM: Human
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (1)..(577)
;; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-81822
```

```
Query Match      0.8%; Score 18; DB 12; Length 577;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1679 AGCCTTCTGATAGGAA 1696
      |||||
DB      385 AGCCTTCTGATAGGAA 402
```

```
RESULT 66
US-10-027-632-81822
;; Sequence 81822, Application US/10027632
;; GENERAL INFORMATION:
;; APPLICANT: Wang, David G.
;; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;; FILE REFERENCE: 108827.129
;; CURRENT APPLICATION NUMBER: US/10/027,632
;; CURRENT FILING DATE: 2002-04-30
;; PRIOR APPLICATION NUMBER: US 60/218,006
;; PRIOR FILING DATE: 2000-07-12
;; PRIOR APPLICATION NUMBER: US 60/198,676
;; PRIOR FILING DATE: 2000-04-20
;; PRIOR APPLICATION NUMBER: US 60/193,483
;; PRIOR FILING DATE: 2000-03-29
;; PRIOR APPLICATION NUMBER: US 60/185,218
;; PRIOR FILING DATE: 2000-02-24
;; PRIOR APPLICATION NUMBER: US 60/167,363
;; PRIOR FILING DATE: 1999-11-23
;; PRIOR APPLICATION NUMBER: US 60/156,358
;; PRIOR FILING DATE: 1999-09-28
;; PRIOR APPLICATION NUMBER: US 60/146,002
;; PRIOR FILING DATE: 1999-08-09
;; NUMBER OF SEQ ID NOS: 325720
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO: 81822
;; LENGTH: 577
;; TYPE: DNA
;; ORGANISM: Human
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (1)..(577)
;; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-81822
```

```
Query Match      0.8%; Score 18; DB 13; Length 577;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1679 AGCCTTCTGATAGGAA 1696
      |||||
DB      385 AGCCTTCTGATAGGAA 402
```

```
RESULT 67
US-10-027-632-185059
;; Sequence 185059, Application US/10027632
;; Publication No. US20030204075A9
;; GENERAL INFORMATION:
;; APPLICANT: Wang, David G.
;; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;; FILE REFERENCE: 108827.129
;; CURRENT APPLICATION NUMBER: US/10/027,632
;; CURRENT FILING DATE: 2002-04-30
;; PRIOR APPLICATION NUMBER: US 60/218,006
;; PRIOR FILING DATE: 2000-07-12
;; PRIOR APPLICATION NUMBER: US 60/198,676
;; PRIOR FILING DATE: 2000-04-20
;; PRIOR APPLICATION NUMBER: US 60/193,483
;; PRIOR FILING DATE: 2000-03-29
;; PRIOR APPLICATION NUMBER: US 60/185,218
;; PRIOR FILING DATE: 2000-02-24
;; PRIOR APPLICATION NUMBER: US 60/167,363
;; PRIOR FILING DATE: 1999-11-23
;; PRIOR APPLICATION NUMBER: US 60/156,358
;; PRIOR FILING DATE: 1999-09-28
;; PRIOR APPLICATION NUMBER: US 60/146,002
;; PRIOR FILING DATE: 1999-08-09
;; NUMBER OF SEQ ID NOS: 325720
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO: 185059
;; LENGTH: 608
;; TYPE: DNA
;; ORGANISM: Human
US-10-027-632-185059
```

```
Query Match      0.8%; Score 18; DB 12; Length 608;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      320 ATCAAGTATGATGAAAA 337
      |||||
DB      183 ATCAAGTATGATGAAAA 200
```

```
RESULT 68
US-10-027-632-185059
;; Sequence 185059, Application US/10027632
;; GENERAL INFORMATION:
;; APPLICANT: Wang, David G.
;; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;; FILE REFERENCE: 108827.129
;; CURRENT APPLICATION NUMBER: US/10/027,632
;; CURRENT FILING DATE: 2002-04-30
;; PRIOR APPLICATION NUMBER: US 60/218,006
;; PRIOR FILING DATE: 2000-07-12
;; PRIOR APPLICATION NUMBER: US 60/198,676
;; PRIOR FILING DATE: 2000-04-20
;; PRIOR APPLICATION NUMBER: US 60/193,483
;; PRIOR FILING DATE: 2000-03-29
;; PRIOR APPLICATION NUMBER: US 60/185,218
;; PRIOR FILING DATE: 2000-02-24
;; PRIOR APPLICATION NUMBER: US 60/167,363
;; PRIOR FILING DATE: 1999-11-23
;; PRIOR APPLICATION NUMBER: US 60/156,358
;; PRIOR FILING DATE: 1999-09-28
;; PRIOR APPLICATION NUMBER: US 60/146,002
;; PRIOR FILING DATE: 1999-08-09
;; NUMBER OF SEQ ID NOS: 325720
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO: 185059
;; LENGTH: 608
;; TYPE: DNA
;; ORGANISM: Human
US-10-027-632-185059
```

Query Match 0.8%; Score 18; DB 13; Length 608;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 320 ATCAAGTGTGATGAAAA 337
|||||
Db 183 ATCAAGTGTGATGAAAA 200

RESULT 69

US-10-027-632-27877/C
; Sequence 27877, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:

; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27877
; LENGTH: 674
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-27877

Query Match 0.8%; Score 18; DB 12; Length 674;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1693 GGAAGGTTGCAGCTCA 1710
|||||
Db 594 GGAAGGTTGCAGCTCA 577

RESULT 70

US-10-027-632-27877/C
; Sequence 27877, Application US/10027632
; GENERAL INFORMATION:

; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27877
; LENGTH: 674
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-27877

RESULT 71

Query Match 0.8%; Score 18; DB 13; Length 674;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1693 GGAAGGTTGCAGCTCA 1710
|||||
Db 594 GGAAGGTTGCAGCTCA 577

US-09-815-242-9810/C
; Sequence 9810, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9810
; LENGTH: 924
; TYPE: DNA
; ORGANISM: Salmonella typhi
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(924)
; NAME/KEY: misc_feature
; LOCATION: (1)...(924)
; OTHER INFORMATION: n = A,T,C or G
US-09-815-242-9810

Query Match 0.8%; Score 18; DB 9; Length 924;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2106 ACTTCAAGCGGATGAAG 2123
|||||
Db 854 ACTTCAAGCGGATGAAG 837

```

RESULT 72
US-09-070-927A-278
; Sequence 278, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunesh
; Patrick J. Dillon
; Steven Barash
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8504
; INFORMATION FOR SEQ ID NO: 278:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1587 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 278:
US-09-070-927A-278

Query Match          0.8%; Score 18; DB 10; Length 1587;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2297 ATGATAACAATAGTATC 2314
Db      410 ATGATAACAATAGTATC 427

RESULT 73
US-10-029-386-24974
; Sequence 24974, Application US/10029386
; Patent No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Hanzel, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20

```

```

; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 24974
; LENGTH: 1619
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO D87009.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.38
; OTHER INFORMATION: EST HUMAN HIT: N39816.1, EVALUE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: Q60821, EVALUE 7.00e-10
; OTHER INFORMATION: NT HIT: g115300639, EVALUE 7.00e-43
US-10-029-386-24974

Query Match          0.8%; Score 18; DB 12; Length 1619;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1111 CTCAAATTCCTTTGGT 1128
Db      1567 CTCAAATTCCTTTGGT 1584

RESULT 74
US-10-029-386-25318
; Sequence 25318, Application US/10029386
; Patent No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Hanzel, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 25318
; LENGTH: 1629
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO D87011.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
; OTHER INFORMATION: SWISSPROT HIT: Q60821, EVALUE 6.00e-09
; OTHER INFORMATION: NT HIT: g115300639, EVALUE 2.00e-43
; OTHER INFORMATION: EST_HUMAN HIT: N39816.1, EVALUE 0.00e+00
US-10-029-386-25318

Query Match          0.8%; Score 18; DB 12; Length 1629;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1111 CTCAAATTCCTTTGGT 1128
Db      1577 CTCAAATTCCTTTGGT 1594

RESULT 75
US-09-833-381-1943/c
; Sequence 1943, Application US/09833381
; Patent No. US20020132090A1
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: NO. US20020132090A1el Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; CURRENT FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSeq for Windows Version 3.0

```

Query Match	0.88; Score 18; DB 14; Length 2956;
-------------	-------------------------------------

Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1905 ATCATTACCATATATTA 1922

Db 317 ATCATTACCATATATTA 300

RESULT 80

US-10-044-090-38

; Sequence 38, Application US/10044090

; Publication No. US20020137081A1

; GENERAL INFORMATION:

; APPLICANT: Olga Bandman

; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION

; FILE REFERENCE: PA-0028 US

; CURRENT APPLICATION NUMBER: US/10/044,090

; NUMBER OF SEQ ID NOS: 850

; SOFTWARE: PERL Program

; SEQ ID NO 38

; LENGTH: 3154

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; OTHER INFORMATION: Inocyte ID No. US20020137081A1 1519105CB1

US-10-044-090-38

Query Match 0.8%; Score 18; DB 13; Length 3154;

Best Local Similarity 100.0%; Pred. No. 2.7e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2195 AAAGAGCAGAAAGTTTG 2212

Db 83 AAAGAGCAGAAAGTTTG 100

RESULT 81

US-10-340-097-5

; Sequence 5, Application US/10340097

; Publication No. US20030162276A1

; GENERAL INFORMATION:

; APPLICANT: Ratner, Amir

; APPLICANT: Sun, Hui

; APPLICANT: Lupski, James R.

; APPLICANT: Nathans, Jeremy

; APPLICANT: Anderson, Kent L.

; APPLICANT: Leppert, Mark

; APPLICANT: Dean, Michael

; APPLICANT: Singh, Nanda

; APPLICANT: Shroyer, No. US20030162276A1h F.

; APPLICANT: Smallwood, Philip M.

; APPLICANT: Allikmets, Rando

; APPLICANT: Lewis, Richard A.

; APPLICANT: Li, Yixin

; TITLE OF INVENTION: Nucleic Acid And Amino Acid Sequences For ATP-Binding Cassette

; TITLE OF INVENTION: Transporter And Methods Of Screening For Agents That Modify ATP-

; FILE REFERENCE: BYLR0065

; CURRENT APPLICATION NUMBER: US/10/340,097

; PRIOR FILING DATE: 2003-01-10

; PRIOR APPLICATION NUMBER: US/09/032,438A

; PRIOR FILING DATE: 1998-02-27

; PRIOR APPLICATION NUMBER: 60/039,388

; NUMBER OF SEQ ID NOS: 120

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 5

; LENGTH: 6705

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-340-097-5

Query Match 0.8%; Score 18; DB 12; Length 6705;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2095 TCCAATATAGAACTTCAA 2112

Db 3699 TCCAATATAGAACTTCAA 3716

RESULT 82

US-10-336-215-5

; Sequence 5, Application US/10336215

; Publication No. US20030170852A1

; GENERAL INFORMATION:

; APPLICANT: Allikmets, Rando

; APPLICANT: Anderson, Kent L.

; APPLICANT: Dean, Michael

; APPLICANT: Leppert, Mark

; APPLICANT: Lewis, Richard A.

; APPLICANT: Li, Yixin

; APPLICANT: Lupski, James R.

; APPLICANT: Nathans, Jeremy

; APPLICANT: Ratner, Amir

; APPLICANT: Shroyer, No. US20030170852A1h F.

; APPLICANT: Singh, Nanda

; APPLICANT: Smallwood, Philip

; APPLICANT: Sun, Hui

; TITLE OF INVENTION: Methods Of Screening And Diagnostics Using ATP-Binding Cassette

; FILE REFERENCE: APPI0089

; CURRENT APPLICATION NUMBER: US/10/336,215

; PRIOR FILING DATE: 2003-04-11

; PRIOR APPLICATION NUMBER: 60/039,388

; PRIOR FILING DATE: 1997-02-27

; PRIOR APPLICATION NUMBER: 09/032,438

; NUMBER OF SEQ ID NOS: 120

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 5

; LENGTH: 6705

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-336-215-5

Query Match 0.8%; Score 18; DB 12; Length 6705;

Best Local Similarity 100.0%; Pred. No. 2.9e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2095 TCCAATATAGAACTTCAA 2112

Db 3699 TCCAATATAGAACTTCAA 3716

RESULT 83

US-10-336-219-5

; Sequence 5, Application US/10336219

; Publication No. US20030170853A1

; GENERAL INFORMATION:

; APPLICANT: Allikmets, Rando

; APPLICANT: Anderson, Kent L.

; APPLICANT: Dean, Michael

; APPLICANT: Leppert, Mark

; APPLICANT: Lewis, Richard A.

; APPLICANT: Li, Yixin

; APPLICANT: Lupski, James R.

; APPLICANT: Nathans, Jeremy

; APPLICANT: Ratner, Amir

; APPLICANT: Shroyer, No. US20030170853A1h F.

; APPLICANT: Singh, Nanda

; APPLICANT: Smallwood, Philip

; APPLICANT: Sun, Hui

; TITLE OF INVENTION: Methods Of Gene Therapy Using Nucleic Acid Sequences For


```

; TITLE OF INVENTION: ATP-Binding Cassette Transporter
; FILE REFERENCE: BYLR0072
; CURRENT APPLICATION NUMBER: US/10/336,219
; CURRENT FILING DATE: 2003-01-03
; PRIOR APPLICATION NUMBER: 60/039,388
; PRIOR FILING DATE: 1997-02-27
; PRIOR APPLICATION NUMBER: 09/032,438
; PRIOR FILING DATE: 1998-02-27
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 5
; LENGTH: 6705
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-336-219-5

Query Match          0.8%; Score 18; DB 12; Length 6705;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      2095  TCCAAATAAGAACTTCAA 2112
Db      3699  TCCAAATAAGAACTTCAA 3716

RESULT 84
US-10-340-097-2
; Sequence 2, Application US/10340097
; Publication No. US20030162276A1
; GENERAL INFORMATION:
; APPLICANT: Ratner, Amir
; APPLICANT: Sun, Hui
; APPLICANT: Lupski, James R.
; APPLICANT: Nathans, Jeremy
; APPLICANT: Anderson, Kent L.
; APPLICANT: Leppert, Mark
; APPLICANT: Dean, Michael
; APPLICANT: Singh, Nanda
; APPLICANT: Shroyer, No. US20030162276A1h F.
; APPLICANT: Smallwood, Philip M.
; APPLICANT: Allikmets, Rando
; APPLICANT: Li, Yixin
; TITLE OF INVENTION: Nucleic Acid And Amino Acid Sequences For ATP-Binding Cassette
; TITLE OF INVENTION: Transporter And Methods Of Screening For Agents That Modify ATP-
; TITLE OF INVENTION: Transporter
; FILE REFERENCE: BYLR0065
; CURRENT APPLICATION NUMBER: US/10/340,097
; CURRENT FILING DATE: 2003-01-10
; PRIOR APPLICATION NUMBER: US/09/032,438A
; PRIOR FILING DATE: 1998-02-27
; PRIOR APPLICATION NUMBER: 60/039,388
; PRIOR FILING DATE: 1997-02-27
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 2
; LENGTH: 6819
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-340-097-2

Query Match          0.8%; Score 18; DB 12; Length 6819;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      2095  TCCAAATAAGAACTTCAA 2112
Db      3699  TCCAAATAAGAACTTCAA 3716

RESULT 85
US-10-336-215-2
; Sequence 2, Application US/10336215
; Publication No. US20030170852A1
; GENERAL INFORMATION:
; APPLICANT: Allikmets, Rando
; APPLICANT: Anderson, Kent L.
; APPLICANT: Dean, Michael
; APPLICANT: Leppert, Mark
; APPLICANT: Lewis, Richard A.
; APPLICANT: Li, Yixin
; APPLICANT: Lupski, James R.
; APPLICANT: Nathans, Jeremy
; APPLICANT: Ratner, Amir
; APPLICANT: Shroyer, No. US20030170852A1h F.
; APPLICANT: Smallwood, Philip
; APPLICANT: Sun, Hui
; TITLE OF INVENTION: Methods Of Screening And Diagnostics Using ATP-Binding Cassette
; TITLE OF INVENTION: Transporter
; FILE REFERENCE: APP10089
; CURRENT APPLICATION NUMBER: US/10/336,215
; CURRENT FILING DATE: 2003-04-11
; PRIOR APPLICATION NUMBER: 60/039,388
; PRIOR FILING DATE: 1997-02-27
; PRIOR APPLICATION NUMBER: 09/032,438
; PRIOR FILING DATE: 1998-02-27
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 2
; LENGTH: 6819
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-336-215-2

Query Match          0.8%; Score 18; DB 12; Length 6819;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      2095  TCCAAATAAGAACTTCAA 2112
Db      3699  TCCAAATAAGAACTTCAA 3716

RESULT 86
US-10-336-219-2
; Sequence 2, Application US/10336219
; Publication No. US20030170853A1
; GENERAL INFORMATION:
; APPLICANT: Allikmets, Rando
; APPLICANT: Anderson, Kent L.
; APPLICANT: Dean, Michael
; APPLICANT: Leppert, Mark
; APPLICANT: Lewis, Richard A.
; APPLICANT: Li, Yixin
; APPLICANT: Lupski, James R.
; APPLICANT: Nathans, Jeremy
; APPLICANT: Ratner, Amir
; APPLICANT: Shroyer, No. US20030170853A1h F.
; APPLICANT: Smallwood, Philip
; APPLICANT: Sun, Hui
; TITLE OF INVENTION: Methods Of Gene Therapy Using Nucleic Acid Sequences For
; TITLE OF INVENTION: ATP-Binding Cassette Transporter
; FILE REFERENCE: BYLR0072
; CURRENT APPLICATION NUMBER: US/10/336,219
; CURRENT FILING DATE: 2003-01-03
; PRIOR APPLICATION NUMBER: 60/039,388
; PRIOR FILING DATE: 1997-02-27
; PRIOR APPLICATION NUMBER: 09/032,438
; PRIOR FILING DATE: 1998-02-27
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 2
; LENGTH: 6819
; TYPE: DNA
```



```
US-09-736-960-86
; Sequence 86, Application US/09736960
; Patent No. US20020102267A1
; GENERAL INFORMATION:
; APPLICANT: Lu, Peter
; APPLICANT: Garman, Jonathan David
; APPLICANT: Candia III, Albert Frederick
; APPLICANT: Arbor Vita Corporation
; TITLE OF INVENTION: CLASP-5 Transmembrane Protein
; FILE REFERENCE: 020054-000511US
; CURRENT APPLICATION NUMBER: US/09736, 960
; CURRENT FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: US 60/160, 860
; PRIOR FILING DATE: 1999-10-21
; PRIOR APPLICATION NUMBER: US 60/162, 498
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: US 60/170, 453
; PRIOR FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: US 60/176, 195
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: US 60/182, 296
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: US 09/547, 276
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196, 267
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196, 460
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196, 527
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196, 528
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 09/687, 837
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240, 503
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240, 508
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240, 539
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240, 543
; NUMBER OF SEQ ID NOS: 134
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 86
; LENGTH: 66686
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: ordered human genomic DNA at CLASP-5 locus
US-09-736-960-86

Query Match          0.8%; Score 18; DB 10; Length 66686;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2144 GGTGAGCCAGAAATCCCT 2161
Db      30662 GGTGAGCCAGAAATCCCT 30679
```

```
RESULT 91
US-10-010-942B-18/c
; Sequence 18, Application US/10010942B
; Publication No. US20030165496A1
; GENERAL INFORMATION:
; APPLICANT: Basi, Gurig
; APPLICANT: Saldanha, Jose
; APPLICANT: Yednock, Ted
; TITLE OF INVENTION: HUMANIZED ANTIBODIES THAT RECOGNIZE
; TITLE OF INVENTION: BETA AMYLOID PEPTIDE
; FILE REFERENCE: ELN-002
; CURRENT APPLICATION NUMBER: US/10/010, 942B
```

```
; CURRENT FILING DATE: 2002-12-06
; PRIOR APPLICATION NUMBER: US 60/251, 892
; PRIOR FILING DATE: 2000-12-06
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 131
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: primer
US-10-010-942B-18

Query Match          0.7%; Score 17; DB 12; Length 131;
Best Local Similarity 100.0%; Pred. No. 6.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      615 TGGCTGCTGCAGAGCC 631
Db      33 TGGCTGCTGCAGAGCC 17
```

```
RESULT 92
US-10-010-942B-19
; Sequence 19, Application US/10010942B
; Publication No. US20030165496A1
; GENERAL INFORMATION:
; APPLICANT: Basi, Gurig
; APPLICANT: Saldanha, Jose
; APPLICANT: Yednock, Ted
; TITLE OF INVENTION: HUMANIZED ANTIBODIES THAT RECOGNIZE
; TITLE OF INVENTION: BETA AMYLOID PEPTIDE
; FILE REFERENCE: ELN-002
; CURRENT APPLICATION NUMBER: US/10/010, 942B
; CURRENT FILING DATE: 2002-12-06
; PRIOR APPLICATION NUMBER: US 60/251, 892
; PRIOR FILING DATE: 2000-12-06
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 146
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: primer
US-10-010-942B-19
```

```
Query Match          0.7%; Score 17; DB 12; Length 146;
Best Local Similarity 100.0%; Pred. No. 6.5e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      615 TGGCTGCTGCAGAGCC 631
Db      9 TGGCTGCTGCAGAGCC 25
```

```
RESULT 93
US-09-878-574-6546/c
; Sequence 6546, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878, 574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333, 535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 6546
```

```

; LENGTH: 271
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 701098572H1
US-09-878-574-6546

Query Match
Best Local Similarity 100.0%; Score 17; DB 10; Length 271;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 92 GAGATTGACTCCTGA 108
Db 67 GAGATTGACTCCTGA 51

RESULT 94
US-09-923-876-4422
; Sequence 4422, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
; APPLICANT: Laligudi, Raghunath V.
; APPLICANT: Kamigaki, Laura Y. (lto)
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
; FILE REFERENCE: PL-0012-1 CON
; CURRENT APPLICATION NUMBER: US/09/923,876
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: 09/298,329
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/085,331
; PRIOR FILING DATE: 1998-05-05
; NUMBER OF SEQ ID NOS: 6332
; SOFTWARE: PERL Program
; SEQ ID NO 4422
; LENGTH: 295
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020013958A1 700455016H1
; NAME/KEY: unsure
; LOCATION: 7, 107, 194, 204, 243, 254, 256, 260, 278, 288, 292-293
; OTHER INFORMATION: a, t, c, g, or other
US-09-923-876-4422

Query Match
Best Local Similarity 100.0%; Score 17; DB 9; Length 295;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1287 AAGAAATGTGCTCCT 1303
Db 212 AAGAAATGTGCTCCT 228

RESULT 95
US-13-106-698-2167/c
; Sequence 2167, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 2167
```

```

; LENGTH: 316
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (7)..(7)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (212)..(212)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (268)..(268)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (276)..(276)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (283)..(284)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (311)..(311)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-2167

Query Match
Best Local Similarity 100.0%; Score 17; DB 14; Length 316;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2116 GGATGAGAGCCACTAG 2132
Db 141 GGATGAGAGCCACTAG 125

RESULT 96
US-09-765-272-149
; Sequence 149, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 149:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 322 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
```

SEQUENCE DESCRIPTION: SEQ ID NO: 149:
US-09-765-272-149

Query Match 0.7%; Score 17; DB 9; Length 322;
Best Local Similarity 100.0%; Pred.No. 7.1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1439 AATAAAGAAAATTGCT 1455
DB 38 AATAAAGAAAATTGCT 54

RESULT 97
US-10-060-036-2151/c
; Sequence 2151, Application US/10060036
; Publication No. US20030073144A1

GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hepler, David H.
; APPLICANT: Hepler, William T.
; APPLICANT: Jiang, Yugu
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
; FILE REFERENCE: 210121.566
; CURRENT APPLICATION NUMBER: US/10/060,036
; CURRENT FILING DATE: 2002-01-30
; NUMBER OF SEQ ID NOS: 4560
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2151
; LENGTH: 322
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-060-036-2151

Query Match 0.7%; Score 17; DB 14; Length 322;
Best Local Similarity 100.0%; Pred.No. 7.1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1690 TAAGCAAAAAGTTGCAG 1706
DB 73 TAAGCAAAAAGTTGCAG 57

RESULT 98
US-09-878-574-3595
; Sequence 3595, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:

; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 3595
; LENGTH: 371
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: LIB3028-008-Q1-B1-C2
US-09-878-574-3595

Query Match 0.7%; Score 17; DB 10; Length 371;
Best Local Similarity 100.0%; Pred.No. 7.2e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 755 AGCAACAAGCAACAC 771

DB 282 AGCAACAAGCAACAC 298

RESULT 99
US-09-757-781-26/c
; Sequence 26, Application US/09757781
; Patent No. US20020137166A1

GENERAL INFORMATION:
; APPLICANT: Reddy, Roopa
; APPLICANT: Tang, Y. Tom
; APPLICANT: Baughn, Mariah R.
; APPLICANT: Krasnow, Randi E.
; TITLE OF INVENTION: ASIP-RELATED PROTEINS
; FILE REFERENCE: PC-0032 US
; CURRENT APPLICATION NUMBER: US/09/757,781
; CURRENT FILING DATE: 2001-01-09
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PERL Program
; SEQ ID NO 26
; LENGTH: 375
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Inocyte ID No. US20020137166A1 5627320R8
US-09-757-781-26

Query Match 0.7%; Score 17; DB 10; Length 375;
Best Local Similarity 100.0%; Pred.No. 7.2e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 AAGATGAGATATTGTT 295
DB 135 AAGATGAGATATTGTT 119

RESULT 100
US-09-918-995-36324

; Sequence 36324, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1998-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 36324
; LENGTH: 402
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-36324

Query Match 0.7%; Score 17; DB 11; Length 402;
Best Local Similarity 100.0%; Pred.No. 7.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1715 TATCTAAAGAAAAGG 1731
DB 386 TATCTAAAGAAAAGG 402

Search completed: November 14, 2003, 10:32:45
Job time : 1716 secs

BEST AVAILABLE COPY

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 10:38:30 ; Search time 462 Seconds
(without alignments)
273.477 Million cell updates/sec

Title: US-09-765-271-56

Perfect score: 796

Sequence: 1 SYELGLYQARTVKNRNVSY.....KLALLKGNPSVSKERIN 796

Scoring table: OLIGO
Gap 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 200 summaries

Database :

A.Geneseq_19Jun03:*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	795	99.9	796	19	AAW55090
2	795	99.9	796	23	ABP54584
3	695	87.3	802	24	ABU01598
4	695	87.3	819	21	AAU01468
5	80	10.1	334	21	AAU12728
6	80	10.1	334	23	AAU84027
7	80	10.1	690	21	AAU12745
8	80	10.1	690	23	AAU84093
9	80	10.1	821	21	AAU12727

10	80	10.1	821	21	AAU12766	Streptococcus pneu
11	80	10.1	821	23	AAU84026	Truncated variant
12	80	10.1	840	21	AAU12716	Streptococcus pneu
13	80	10.1	840	23	AAU75933	Streptococcus pneu
14	60	7.5	805	21	AAU12764	Streptococcus pneu
15	60	7.5	807	21	AAU12765	Streptococcus pneu
16	60	7.5	811	21	AAU12760	Streptococcus pneu
17	60	7.5	811	21	AAU12761	Streptococcus pneu
18	60	7.5	811	21	AAU12762	Streptococcus pneu
19	60	7.5	811	21	AAU12763	Streptococcus pneu
20	60	7.5	816	21	AAU12766	Streptococcus pneu
21	60	7.5	816	21	AAU12757	Streptococcus pneu
22	60	7.5	816	21	AAU12758	Streptococcus pneu
23	60	7.5	819	21	AAU12740	Streptococcus pneu
24	60	7.5	819	21	AAU12754	Streptococcus pneu
25	60	7.5	819	21	AAU01469	Recombinant varian
26	60	7.5	819	23	AAU84087	Truncated variant
27	60	7.5	819	24	AAU01597	S. pneumoniae type
28	60	7.5	820	21	AAU12755	Streptococcus pneu
29	60	7.5	826	21	AAU91939	S. pneumoniae 92 k
30	60	7.5	827	21	AAU81662	Streptococcus pneu
31	60	7.5	834	21	AAU12759	Streptococcus pneu
32	60	7.5	838	21	AAU12720	Streptococcus pneu
33	60	7.5	838	21	AAU01466	Recombinant varian
34	60	7.5	838	23	AAU75934	Streptococcus pneu
35	60	7.5	839	24	ABU01418	S. pneumoniae type
36	56	7.0	613	21	AAU12730	Streptococcus pneu
37	56	7.0	613	23	AAU84029	Truncated variant
38	51	6.4	721	20	AAU05753	Streptococcus pneu
39	51	6.4	763	19	AAU55095	Streptococcus pneu
40	51	6.4	763	23	AAU54589	S. pneumoniae SP04
41	32	4.0	42	23	AAU83828	S. pneumoniae anti
42	32	4.0	205	23	AAU12726	Streptococcus pneu
43	32	4.0	205	23	AAU84025	Truncated variant
44	32	4.0	447	19	AAU61228	Streptococcus pneu
45	32	4.0	447	23	ABP54647	S. pneumoniae SP10
46	32	4.0	484	21	AAU12718	Streptococcus pneu
47	32	4.0	484	21	AAU01467	Recombinant varian
48	32	4.0	484	21	AAU81708	Streptococcus pneu
49	32	4.0	485	21	AAU81538	Streptococcus pneu
50	32	4.0	489	21	AAU12723	Streptococcus pneu
51	32	4.0	489	23	AAU84022	Truncated variant
52	32	4.0	509	21	AAU12724	Streptococcus pneu
53	32	4.0	509	23	AAU84023	Streptococcus pneu
54	32	4.0	679	23	AAU84092	Truncated variant
55	32	4.0	780	21	AAU12744	Streptococcus pneu
56	32	4.0	840	21	AAU12721	Streptococcus pneu
57	32	4.0	840	23	AAU76151	Streptococcus pneu
58	32	4.0	1019	21	AAU12722	Streptococcus pneu
59	32	4.0	1019	21	AAU12748	Streptococcus pneu
60	32	4.0	1019	21	AAU12749	Streptococcus pneu
61	32	4.0	1019	21	AAU12750	Streptococcus pneu
62	32	4.0	1019	21	AAU12751	Streptococcus pneu
63	32	4.0	1019	21	AAU12752	Streptococcus pneu
64	32	4.0	1019	21	AAU12753	Streptococcus pneu
65	32	4.0	1019	23	AAU84021	Truncated variant
66	32	4.0	1039	21	AAU12715	Streptococcus pneu
67	32	4.0	1039	23	AAU75932	Streptococcus pneu
68	32	4.0	1039	24	AAU01419	S. pneumoniae type
69	27	3.4	294	23	AAU84094	Truncated variant
70	27	3.4	466	23	ABP26208	Streptococcus poly
71	27	3.4	473	21	AAU12743	Streptococcus pneu
72	27	3.4	473	23	AAU84091	Truncated variant
73	27	3.4	568	21	AAU12741	Streptococcus pneu
74	27	3.4	568	23	AAU84088	Truncated variant
75	27	3.4	612	23	AAU84031	Truncated variant
76	27	3.4	1139	23	AAU84055	S. pneumoniae deri
77	27	3.4	1378	23	AAU84053	S. pneumoniae 20 k
78	25	3.1	168	21	AAU91938	Streptococcus pneu
79	24	3.0	226	21	AAU12738	Truncated variant
80	24	3.0	226	23	AAU84085	Streptococcus pneu
81	24	3.0	428	21	AAU12735	Truncated variant
82	24	3.0	428	23	AAU84048	Truncated variant

83	24	3.0	555	21	AA812734	Streptococcus pneu	156	7	0.9	27	21	AA814205	Human Lyn A kinase
84	24	3.0	555	23	AA84047	Truncated variant	157	7	0.9	27	22	AA814631	Human tyrosine kin
85	24	3.0	999	23	AA84051	S. pneumoniae deri	158	7	0.9	34	22	AA898647	N-terminal peptide
86	24	3.0	999	23	AA84052	S. pneumoniae deri	159	7	0.9	41	21	AA898389	Streptococcus pneu
87	24	3.0	1126	23	AA84058	S. pneumoniae deri	160	7	0.9	45	21	AA814203	Human Lyn B kinase
88	24	3.0	1238	23	AA84056	S. pneumoniae deri	161	7	0.9	45	23	AA814629	Human tyrosine kin
89	24	3.0	1365	23	AA84057	S. pneumoniae deri	162	7	0.9	45	22	AA865490	Human immune/haema
90	24	2.9	94	23	AA835827	S. pneumoniae anti	163	7	0.9	47	22	AA826643	Streptococcus poly
91	17	2.1	163	20	AA85752	Streptococcus pneu	164	7	0.9	49	22	AA822724	Protein #4723 enco
92	17	2.1	381	21	AA891286	Group B Streptococ	165	7	0.9	49	23	AA840208	Human peptidase
93	17	2.1	793	20	AA827347	Group B Streptococ	166	7	0.9	50	18	AA845409	N-terminal amino a
94	17	2.1	822	21	AA891283	Group B Streptococ	167	7	0.9	50	23	AA804572	Human ORFX protein
95	17	2.1	822	22	AA800028	Streptococcus agal	168	7	0.9	53	23	AA834598	Human ORF571 prot
96	17	2.1	822	22	AA800030	Streptococcus agal	169	7	0.9	63	22	AA802094	Human polypeptide
97	17	2.1	822	23	AA828904	Streptococcus poly	170	7	0.9	66	21	AA814202	Human Lyn A kinase
98	17	2.1	822	23	AA829703	Streptococcus poly	171	7	0.9	66	23	AA814628	Human tyrosine kin
99	17	2.1	823	21	AA812746	Streptococcus pneu	172	7	0.9	66	23	AA813888	Human ORF1861 prot
100	17	2.1	824	21	AA812747	Streptococcus pneu	173	7	0.9	70	21	AA802963	Human secreted pro
101	17	2.1	825	22	AA800026	Streptococcus pyog	174	7	0.9	76	21	AA844881	Human secreted pro
102	17	2.1	825	22	AA800029	Streptococcus pyog	175	7	0.9	76	21	AA844882	Human secreted pro
103	17	2.1	825	23	AA826207	Streptococcus poly	176	7	0.9	81	22	AA874664	Human colon cancer
104	15	1.9	1057	21	AA812725	Streptococcus pneu	177	7	0.9	85	22	AA808815	Human polypeptide
105	15	1.9	1058	23	AA840097	S. pneumoniae deri	178	7	0.9	86	21	AA832570	Eucalyptus grandis
106	14	1.8	487	21	AA812729	Streptococcus pneu	179	7	0.9	91	22	AA895891	Human testicular a
107	14	1.8	487	21	AA844028	Truncated variant	180	7	0.9	91	22	AA895192	Human reproductive
108	14	1.8	487	21	AA844090	Truncated variant	181	7	0.9	94	21	AA833225	Human neoplasia
109	13	1.6	132	21	AA812737	Streptococcus pneu	182	7	0.9	97	22	AA821570	Novel human neopla
110	13	1.6	132	23	AA840050	Truncated variant	183	7	0.9	97	22	AA821740	Novel human neopla
111	12	1.5	203	21	AA812739	Streptococcus pneu	184	7	0.9	99	11	AA805700	MP-57 protein, M
112	12	1.5	203	23	AA840086	Truncated variant	185	7	0.9	99	23	AA842619	Human ovarian anti
113	12	1.5	214	23	AA840095	Truncated variant	186	7	0.9	100	22	AA881221	Mycobacterium tube
114	12	1.5	342	23	AA840430	Truncated variant	187	7	0.9	100	22	AA824631	Mycobacterium tube
115	12	1.5	715	20	AA827348	Group B Streptococ	188	7	0.9	100	23	AA850739	Mycobacterium tube
116	12	1.5	807	23	AA833997	Truncated variant	189	7	0.9	103	22	AA817147	Human nervous syst
117	12	1.5	913	23	AA840059	S. pneumoniae deri	190	7	0.9	110	23	AA855485	Lactococcus lactis
118	12	1.5	1152	23	AA840054	S. pneumoniae deri	191	7	0.9	115	22	AA811188	Streptococcus epi
119	11	1.4	485	21	AA891310	Group B Streptococ	192	7	0.9	117	23	AA839740	Streptococcus epi
120	10	1.3	189	24	ABU01420	S. pneumoniae type	193	7	0.9	129	22	AA808365	Novel human diagno
121	9	1.1	121	21	AA812736	Streptococcus pneu	194	7	0.9	145	22	AA805153	Novel human diagno
122	9	1.1	121	23	AA840409	Truncated variant	195	7	0.9	145	23	AA848072	Listeria monocytog
123	9	1.1	272	23	AA84040	Variant of S. pneu	196	7	0.9	154	22	ABG10403	Novel human diagno
124	9	1.1	921	15	AA856487	TAR-binding prote	197	7	0.9	157	21	AA832303	Catalase acid triac
125	9	1.1	921	17	AA806077	Drosophila TARA-B1	198	7	0.9	158	20	AA878484	S.aureus grea prot
126	9	1.1	921	18	AA825028	TARA-Binding prote	199	7	0.9	158	23	AA870181	Human prey protein
127	9	1.1	921	22	AB861528	Drosophila melanog	200	7	0.9	166	22	AA892915	C glutamicum prote
128	9	1.1	921	22	AB866055	S. pneumoniae anti							
129	8	1.0	86	23	AA893838	Variant of S. pneu							
130	8	1.0	270	23	AA84042	Streptococcus pneu							
131	8	1.0	272	23	AA876153	Variant of S. pneu							
132	8	1.0	272	23	AA850937	Helicobacter pylor							
133	8	1.0	323	23	ABU50997	Streptococcus poly							
134	8	1.0	379	23	ABP26206	Helicobacter pylor							
135	8	1.0	481	23	ABU51004	Staphylococcus aur							
136	8	1.0	482	22	AAU33905	Staphylococcus aur							
137	8	1.0	488	22	AAU36971	Staphylococcus epi							
138	8	1.0	490	23	ABP40271	Modified Interfero							
139	8	1.0	503	22	AA815103	Human intracellular							
140	8	1.0	503	22	AAE09329	S. pneumoniae deri							
141	8	1.0	895	23	AA84069	S. pneumoniae deri							
142	8	1.0	896	23	AA84077	S. pneumoniae deri							
143	8	1.0	901	23	AA84068	S. pneumoniae deri							
144	8	1.0	902	23	AA84076	S. pneumoniae deri							
145	8	1.0	907	23	AA84060	S. pneumoniae deri							
146	8	1.0	907	23	AA84061	S. pneumoniae deri							
147	8	1.0	907	23	AA84062	S. pneumoniae deri							
148	8	1.0	907	23	AA84063	S. pneumoniae deri							
149	8	1.0	907	23	AA84064	S. pneumoniae deri							
150	8	1.0	907	23	AA84065	S. pneumoniae deri							
151	8	1.0	1104	24	ABJ25604	Aspergillus fumiga							
152	8	1.0	1289	22	AA881208	Mycobacterium tube							
153	8	1.0	1353	24	ABJ26204	Aspergillus fumiga							
154	8	1.0	1659	24	ABP81332	Streptococcus pneu							
155	8	1.0	1659	24	ABU00861	S. pneumoniae type							

RESULT 1																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																															
----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

XX 30-OCT-1997; 97WO-US19422.
 XX
 XX 31-OCT-1996; 96US-0029960.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Choi GH, Hromockyj A, Johnson LS, Kunsch CA;
 XX
 XX WPI; 1998-272224/24.
 XX N-PSDB; AAV27351.
 XX
 XX Nucleic acid encoding antigenic peptide(s) from Streptococcus
 XX pneumoniae - or their epitope-containing fragments, useful in
 XX protective or therapeutic vaccines, and for diagnosis
 XX
 XX Claim 11; Page 59-60; 118pp; English.
 XX
 XX The present sequence represents a protein from Streptococcus pneumoniae.
 XX The nucleic acid sequence encoding the Streptococcus pneumoniae protein
 XX can be useful in vaccines for inducing protective antibodies against
 XX Streptococcus pneumoniae, for treatment or prevention of infection e.g.
 XX pneumonia, otitis media or meningitis. Probes based on the nucleic acid
 XX are used to detect Streptococcus infection (by usual hybridisation or
 XX amplification methods), also for isolating Streptococcus genes or their
 XX allelic variants. The protein can be used similarly to detect specific
 XX antibodies in standard immunoassays, especially for diagnosing or
 XX monitoring infections. Antibodies which bind the protein are used to
 XX detect corresponding antigens, to purify the protein and for passive
 XX immunisation (optionally coupled to a toxin). Vaccines are administered,
 XX e.g. by injection, orally or through the skin, typically at 0.01-1000
 XX (especially 10-300) mu g/ml per dose.

Sequence 796 AA:

Query Match 99.9%; Score 795; DB 19; Length 796;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYELGIYQARTYKENVNRSVITDQKQNTENTLTPREVSRBEINAEQIVIKITDGGYVT 60
 DB 1 SYELGIYQARTYKENVNRSVITDQKQNTENTLTPREVSRBEINAEQIVIKITDGGYVT 60
 QY 61 SHGDHYHYNGKVPYDAIISEBLMKDPYKLEDEIVNEVKGYIYKDGKYVYLKDA 120
 DB 61 SHGDHYHYNGKVPYDAIISEBLMKDPYKLEDEIVNEVKGYIYKDGKYVYLKDA 120
 QY 61 SHGDHYHYNGKVPYDAIISEBLMKDPYKLEDEIVNEVKGYIYKDGKYVYLKDA 120
 DB 61 SHGDHYHYNGKVPYDAIISEBLMKDPYKLEDEIVNEVKGYIYKDGKYVYLKDA 120
 QY 121 AAADVNTKEEINRQKQESHQREGGT PRNDGAVALARSGRYTTDDGYI FNASDIIEPT 180
 DB 121 AAADVNTKEEINRQKQESHQREGGT PRNDGAVALARSGRYTTDDGYI FNASDIIEPT 180
 QY 121 AAADVNTKEEINRQKQESHQREGGT PRNDGAVALARSGRYTTDDGYI FNASDIIEPT 180
 DB 121 AAADVNTKEEINRQKQESHQREGGT PRNDGAVALARSGRYTTDDGYI FNASDIIEPT 180
 QY 181 GDAYIVPHGDHYHYTKNELSASELAFAEFLSGRGLNSRTRYRQNSDNTSRTWVWS 240
 DB 181 GDAYIVPHGDHYHYTKNELSASELAFAEFLSGRGLNSRTRYRQNSDNTSRTWVWS 240
 QY 181 GDAYIVPHGDHYHYTKNELSASELAFAEFLSGRGLNSRTRYRQNSDNTSRTWVWS 240
 DB 181 GDAYIVPHGDHYHYTKNELSASELAFAEFLSGRGLNSRTRYRQNSDNTSRTWVWS 240
 QY 241 VSNPGTTNTNTSNSTNSQASQNDISLKLQYLPLSQRVESDGLVFDPAQITSTR 300
 DB 241 VSNPGTTNTNTSNSTNSQASQNDISLKLQYLPLSQRVESDGLVFDPAQITSTR 300
 QY 241 VSNPGTTNTNTSNSTNSQASQNDISLKLQYLPLSQRVESDGLVFDPAQITSTR 300
 DB 241 VSNPGTTNTNTSNSTNSQASQNDISLKLQYLPLSQRVESDGLVFDPAQITSTR 300
 QY 301 ARGVAVPHGDHYHYFISQWSELEERLARIIPLRYSNHWVPDRPEQSPQPTPEPSRQ 360
 DB 301 ARGVAVPHGDHYHYFISQWSELEERLARIIPLRYSNHWVPDRPEQSPQPTPEPSRQ 360
 QY 301 ARGVAVPHGDHYHYFISQWSELEERLARIIPLRYSNHWVPDRPEQSPQPTPEPSRQ 360
 DB 301 ARGVAVPHGDHYHYFISQWSELEERLARIIPLRYSNHWVPDRPEQSPQPTPEPSRQ 360
 QY 361 POPAPMLKIDSNSLSQVLRKGEVYBEKGISRYVFAKDLPSSETVKNLESKLSKQBS 420
 DB 361 POPAPMLKIDSNSLSQVLRKGEVYBEKGISRYVFAKDLPSSETVKNLESKLSKQBS 420
 QY 361 POPAPMLKIDSNSLSQVLRKGEVYBEKGISRYVFAKDLPSSETVKNLESKLSKQBS 420
 DB 361 POPAPMLKIDSNSLSQVLRKGEVYBEKGISRYVFAKDLPSSETVKNLESKLSKQBS 420
 QY 421 VSHITLPAKKENVA PROEFYDKAYNLTLEAHKALFNKKGNSPFOALDKLELBNESNT 480
 DB 421 VSHITLPAKKENVA PROEFYDKAYNLTLEAHKALFNKKGNSPFOALDKLELBNESNT 480
 QY 421 VSHITLPAKKENVA PROEFYDKAYNLTLEAHKALFNKKGNSPFOALDKLELBNESNT 480
 DB 421 VSHITLPAKKENVA PROEFYDKAYNLTLEAHKALFNKKGNSPFOALDKLELBNESNT 480
 QY 481 KEKLVDDLAFAPITHPERLGRKNSQIEYTEDEVAIAQLADKYTTSDDGYIPDEHDIISD 540
 DB 481 KEKLVDDLAFAPITHPERLGRKNSQIEYTEDEVAIAQLADKYTTSDDGYIPDEHDIISD 540

DB 481 KEKLVDDLAFAPITHPERLGRKNSQIEYTEDEVAIAQLADKYTTSDDGYIPDEHDIISD 540
 QY 541 EGDAYVTPMGSNHWIGKDSLSDKEKVAQAATYKEXGILPPSPDADVKANPTGDSAAAY 600
 DB 541 EGDAYVTPMGSNHWIGKDSLSDKEKVAQAATYKEXGILPPSPDADVKANPTGDSAAAY 600
 QY 601 NRKGEKRIPLVRLPYWEHTVEVKNGLIIPKHQYHNKIKFAPEDDHTYKAPNGYTLSD 660
 DB 601 NRKGEKRIPLVRLPYWEHTVEVKNGLIIPKHQYHNKIKFAPEDDHTYKAPNGYTLSD 660
 QY 661 LPATIKYIYEHDEPHSDNGWGNASEHYLGKKHSEDPNKNFKADEEVEESTPAPEVP 720
 DB 661 LPATIKYIYEHDEPHSDNGWGNASEHYLGKKHSEDPNKNFKADEEVEESTPAPEVP 720
 QY 721 QVETEKVEAQLKEAEVYLAKVTDSILKANATEETLAGLRNLTLOIWDNNSIIMAEKELTA 780
 DB 721 QVETEKVEAQLKEAEVYLAKVTDSILKANATEETLAGLRNLTLOIWDNNSIIMAEKELTA 780
 QY 781 LKGSNPSVSKEKIN 796
 DB 781 LKGSNPSVSKEKIN 796
 RESULT 2
 ID ABP54584 standard; Protein; 796 AA.
 XX
 AC ABP54584;
 XX
 DT 04-SEP-2002 (first entry)
 XX
 DE S. pneumoniae SP036 protein sequence SEQ ID NO:56.
 XX
 KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
 KW antibacterial; Streptococcal infection; detection.
 XX
 PN Streptococcus pneumoniae.
 XX
 PF US2002061545-A1.
 XX
 PD 23-MAY-2002.
 XX
 PF 22-JAN-2001; 2001US-0765272.
 XX
 PR 30-OCT-1997; 97US-0961083.
 XX
 PA (CHOI/) CHOI G H.
 PA (KUNS/) KUNSCH C A.
 PA (BARA/) BARASH S C.
 PA (DILL/) DILLON P J.
 PA (DOUG/) DOUGHERTY B.
 PA (FANN/) FANNON M R.
 PA (ROSE/) ROSEN C A.
 XX
 PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
 PI Rosen CA;
 XX
 XX WPI; 2002-479261/51.
 XX N-PSDB; AB084819.
 XX
 PT New Streptococcus pneumoniae antigens, useful for detecting
 PT Streptococcus and for preventing or attenuating disease caused by
 PT Streptococcus infection -
 XX
 PS Claim 11; Page 27; 70pp; English.
 XX
 CC AB084792 to AB084904 represents nucleic acids which encode the
 CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669.
 CC The S. pneumoniae antigens have antibacterial activity and can be
 CC used in vaccines. The S. pneumoniae antigens can also be used to
 CC prevent or attenuate a Streptococcal infection in an animal. The
 CC polynucleotides encoding the S. pneumoniae antigens can be used to
 CC detect Streptococcus nucleic acids. AB084905 to AB085130 represent

CC primers used in the cloning of *S. pneumoniae* ORFs (open reading frames)
 CC which are used in an example from the present invention.

XX Sequence 796 AA;
 CC Query Match 99.9%; Score 795; DB 23; Length 796;
 CC Best Local Similarity 100.0%; Pred. No. 0;
 CC Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SYELGLYQARTYKNNRVSITDGKATOKTENLTPDEVSKREGINAEOIYIKITDQGYVT 60
 DB 1 SYELGLYQARTYKNNRVSITDGKATOKTENLTPDEVSKREGINAEOIYIKITDQGYVT 60
 QY 61 SHGDHYHYNGKVPYDAIIESEELMKDPYKDKEDINEVGVGVYIKVDGKYYYLLDA 120
 DB 61 SHGDHYHYNGKVPYDAIIESEELMKDPYKDKEDINEVGVGVYIKVDGKYYYLLDA 120
 QY 121 AHADNVRTKEEINRQKQESQREGEPTPRNDGAVALARASQGRYTTDDGYIFNASDIEDT 180
 DB 121 AHADNVRTKEEINRQKQESQREGEPTPRNDGAVALARASQGRYTTDDGYIFNASDIEDT 180
 QY 181 GDAYIVPRGDHYHYIPKNELSASELAABAFISGNGINSNRTYRONSNTSRTNWPS 240
 DB 181 GDAYIVPRGDHYHYIPKNELSASELAABAFISGNGINSNRTYRONSNTSRTNWPS 240
 QY 241 VSNPGTNTNTNSNTNSQASQSDIDSLKQYKLPLOSQHVSDGLVDPDAQITRT 300
 DB 241 VSNPGTNTNTNSNTNSQASQSDIDSLKQYKLPLOSQHVSDGLVDPDAQITRT 300
 QY 301 ARGVAVPRGDHYHYIPYSQMSLEERIRIIPLYRSNNHWVDSRPEQSPQPTPEPSPG 360
 DB 301 ARGVAVPRGDHYHYIPYSQMSLEERIRIIPLYRSNNHWVDSRPEQSPQPTPEPSPG 360
 QY 361 POPANLKLIDNSSLSVQLVRVGVGVYFEEKIGIRYFADLPBETVYNLESKSKSES 420
 DB 361 POPANLKLIDNSSLSVQLVRVGVGVYFEEKIGIRYFADLPBETVYNLESKSKSES 420
 QY 421 VSHITLAKENAVPRDQEFYDKAYMLTEAHKALFXNKGNSDFQALDKLERLNDSTN 480
 DB 421 VSHITLAKENAVPRDQEFYDKAYMLTEAHKALFXNKGNSDFQALDKLERLNDSTN 480
 QY 481 KEKLVDDLAFAPITHEBERLQKPSQIETDEVRIAQLADKYTTSQGYTFDEHDIISD 540
 DB 481 KEKLVDDLAFAPITHEBERLQKPSQIETDEVRIAQLADKYTTSQGYTFDEHDIISD 540
 QY 541 EGDAYVTPHMGSHWIGDLSDEKKAQAQATYKKGILPPSPDADVQANPTGDSAAIY 600
 DB 541 EGDAYVTPHMGSHWIGDLSDEKKAQAQATYKKGILPPSPDADVQANPTGDSAAIY 600
 QY 601 NRVGGEKRIPLVRLPYWEHTVEVNGNLIIPHKHYNINIKFAMFDDHTYKAPNGYTLSD 660
 DB 601 NRVGGEKRIPLVRLPYWEHTVEVNGNLIIPHKHYNINIKFAMFDDHTYKAPNGYTLSD 660
 QY 661 LFAITIKYVEHPDERPHSNDGMASEHVILGKQHSDEPNKNFKADEEVEETPAPEEVP 720
 DB 661 LFAITIKYVEHPDERPHSNDGMASEHVILGKQHSDEPNKNFKADEEVEETPAPEEVP 720
 QY 721 QVEFEKVAQKKEAEVLLAKYTDSSLKANATETLGLNNNTLTQIMDNNSIIMAEKELLA 780
 DB 721 QVEFEKVAQKKEAEVLLAKYTDSSLKANATETLGLNNNTLTQIMDNNSIIMAEKELLA 780
 QY 781 LKGSNPSVSKEKIN 796
 DB 781 LKGSNPSVSKEKIN 796

RESULT 3
 ABU01598
 ID ABU01598 standard; Protein; 802 AA.

XX ABU01598;
 XX 11-FEB-2003 (first entry)

XX DE S. pneumoniae type 4 strain protein from coding region #1174.
 XX Bacterial meningitis; pneumonia; sepsis; otitis media;
 XX ear infection; antiinflammatory; antibacterial; immunostimulant;
 XX auditory; respiratory; gene therapy; vaccine.
 OS Streptococcus pneumoniae type 4 strain.
 XX WO200277021-A2.
 XX 03-OCT-2002.
 XX 27-MAR-2002; 2002WO-1B02163.
 XX 27-MAR-2001; 2001GB-0007658.
 XX (CHIR-) CHIRON SPA.
 XX (GENO-) INST GENOMIC RES.
 XX Maignani V, Tettelin H, Fraser C;
 XX WPI; 2003-040579/03.
 XX N-PSDB; ABX06886.
 PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,
 PT useful as medicaments for treating a disease or infection
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
 PT or ear infection
 PS Claim 1; SEQ ID No 2348; 56bp; English.
 CC The invention relates to a protein comprising or having at least 50%
 CC identity to any of the 2469 amino acid sequences, identified in the
 CC specification (available on a computer readable format), or its fragment,
 CC expressed from 2469 of 2489 identified DNA coding regions from the
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
 CC AB556654. Also included are an antibody which binds one of the
 CC proteins, treating a patient by administering the protein, DNA or
 CC antibody (in a composition), a kit comprising first and second primers,
 CC which are the nucleic acid cited above or fragments between nucleotides
 CC 8-100 of a sequence not defined in the specification, for amplifying a
 CC target sequence contained within a Streptococcus nucleic acid sequence,
 CC where the first primer is substantially complementary to the target
 CC sequence and the second primer is substantially complementary to the
 CC complement of the target sequence, and where the pairs of the primers
 CC having substantial complementarity define the termini of the target
 CC sequence to be amplified, assay comprising contacting a test compound
 CC with the protein, and determining whether the test compound binds to the
 CC protein and a Streptococcus pneumoniae bacterium, where one or more
 CC genes encoding the proteins has been rendered inactive. The proteins,
 CC nucleic acid molecules, antibody and compositions are useful as
 CC medicaments for treating or preventing a disease or infection due to
 CC streptococcus bacteria, particularly *S. pneumoniae*, such as pneumonia,
 CC sepsis, otitis media or ear infection. They are also useful in developing
 CC vaccines, diagnostics and antibiotics. The methods are useful for
 CC identifying immunodominant proteins. The present sequence is one of
 CC the 2469 proteins expressed by the identified coding regions from the
 CC genomic sequence.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WPI at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 802 AA;
 CC Query Match 87.3%; Score 695; DB 24; Length 802;
 CC Best Local Similarity 99.9%; Pred. No. 0;
 CC Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 SYELGLYQARTYKNNRVSITDGKATOKTENLTPDEVSKREGINAEOIYIKITDQGYVT 60
 DB 7 SYELGLYQARTYKNNRVSITDGKATOKTENLTPDEVSKREGINAEOIYIKITDQGYVT 66

```

QY 61 SHGDHYHYNGKVPYDAIISEELMKDPNYKLEDEIVNEVGGVYIKVDGKYVYLKDA 120
|
|
|
Db 67 SHGDHYHYNGKVPYDAIISEELMKDPNYKLEDEIVNEVGGVYIKVDGKYVYLKDA 126
|
|
|
QY 121 AHADVNRTEEINRQKQESHQREGGTFRNDGAVALARSGRYTTDDGYIFNASDIIEBT 180
|
|
|
Db 127 AHADVNRTEEINRQKQESHQREGGTFRNDGAVALARSGRYTTDDGYIFNASDIIEBT 186
|
|
|
QY 181 GDAYIVPHGDHYHYPKNELSASELAAPFLSGRGNLSNSTRYRQNSDNTSRTNWVS 240
|
|
|
Db 187 GDAYIVPHGDHYHYPKNELSASELAAPFLSGRGNLSNSTRYRQNSDNTSRTNWVS 246
|
|
|
QY 241 VSNPGTTNTNTNSNTNSQASQNSIDSLKQYKLPQSQRHVESDGLVPDPAQITST 300
|
|
|
Db 247 VSNPGTTNTNTNSNTNSQASQNSIDSLKQYKLPQSQRHVESDGLVPDPAQITST 306
|
|
|
QY 301 ARGVAVPHGDHYHYPYSQMSLEERIANIIPLRYSNHWVDSRREPSPQPTPEPSPG 360
|
|
|
Db 307 ARGVAVPHGDHYHYPYSQMSLEERIANIIPLRYSNHWVDSRREPSPQPTPEPSPG 366
|
|
|
QY 361 PQPAPNLKIDSNSLVSQLVRKVGEGYVEEKGISRYFAKDLPESTVKNLESKQES 420
|
|
|
Db 367 PQPAPNLKIDSNSLVSQLVRKVGEGYVEEKGISRYFAKDLPESTVKNLESKQES 426
|
|
|
QY 421 VSHITLAKKENYAPRDQEFYDKAYNLLTEAHKALFNKGRNSDFOLDKLERLNDESTN 480
|
|
|
Db 427 VSHITLAKKENYAPRDQEFYDKAYNLLTEAHKALFNKGRNSDFOLDKLERLNDESTN 486
|
|
|
QY 481 KEKLVDDLAFAPITHPERLQKPSQIEYTEDEVRIAQLADKYTTSDGYIFDEHDIID 540
|
|
|
Db 487 KEKLVDDLAFAPITHPERLQKPSQIEYTEDEVRIAQLADKYTTSDGYIFDEHDIID 546
|
|
|
QY 541 EEDDAVYTPMGSHWTKGDSISPKENVAQAATYKEGIIIPSPDADYKANPTGDSAAAY 600
|
|
|
Db 547 EEDDAVYTPMGSHWTKGDSISPKENVAQAATYKEGIIIPSPDADYKANPTGDSAAAY 606
|
|
|
QY 601 NRKGEKRIPLVRLPVMVEHTEVKNKNLIIPKHDIYHNIKPFMPDHTYYKAPNGTYLTD 660
|
|
|
Db 607 NRKGEKRIPLVRLPVMVEHTEVKNKNLIIPKHDIYHNIKPFMPDHTYYKAPNGTYLTD 666
|
|
|
QY 661 LEFATIKYVEHPDERHSNDGNGNASEHVLGKKHSDENKQFKADEEPVEETPAPEVP 720
|
|
|
Db 667 LEFATIKYVEHPDERHSNDGNGNASEHVLGKKHSDENKQFKADEEPVEETPAPEVP 726
|
|
|
QY 721 QVETEKVEQAQEAETLAKVTDSISKANATETLAGLRNNLLQIWDNNSIMAEKILIA 780
|
|
|
Db 727 QVETEKVEQAQEAETLAKVTDSISKANATETLAGLRNNLLQIWDNNSIMAEKILIA 786
|
|
|
QY 781 LKGSNPSGSKEKIN 796
|
|
|
Db 787 LKGSNPSGSKEKIN 802
|
|
|

```

RESULT 4

```

AAB01468
ID AAB01468 strandcd; Protein; 819 AA.
XX
AC AAB01468;
XX
DT 20-OCT-2000 (first entry)
XX
DE Recombinant variant of Sp36 (Sp36A) of S. pneumoniae.
XX
KW Streptococcus pneumoniae; infection; vaccine; coiled coil region;
KW histidine triad residue; Sp36; antibody; otitis media;
KW nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
KW meningitis; lobar pneumonia.
XX
OS Streptococcus pneumoniae.
XX
FH Key Location/Qualifiers
FT Region 63..68

```

```

FT FT
FT Region /label= Histidine triad residue
FT 118..145
FT /label= Coiled coil region
FT 189..194
FT Region
FT /label= Histidine triad residue
FT 309..314
FT Region
FT /label= Histidine triad residue
FT 406..434
FT Region
FT /label= Coiled coil region
FT 462..493
FT Region
FT /label= Coiled coil region
FT 550..555
FT Region
FT /label= Histidine triad residue
FT 634..639
FT Region
FT /label= Histidine triad residue
FT 724..751
FT /label= Coiled coil region
FT
FT WO20037105-A2.
FT
FT 29-JUN-2000.
FT
FT 21-DEC-1999; 99WO-US30390.
FT
FT 21-DEC-1998; 98US-0113048.
FT
FT (MEDI-) MEDIMUNE INC.
FT
FT Johnson LS, Koenig S, Adamou JE;
FT WPI; 2000-452129/39.
FT N-PSDB; AAA47604.
FT
FT Vaccine useful for prophylaxis and treatment of pneumococcal infections
FT such as otitis media, nasopharyngeal and bronchial infections,
FT comprises Streptococcus pneumoniae proteins
FT
FT Claim 1; Page 61-64; 70pp; English.
FT
FT Although a number of proteins have been suggested as being involved
FT in the pathogenicity of Streptococcus pneumoniae, there still remains
FT a need to identify polypeptides having epitopes in common from
FT various strains of S. pneumoniae in order to utilize such
FT polypeptides in vaccines to protect against a wide variety of
FT S. pneumoniae. New vaccine compositions are described which comprise a
FT Streptococcus pneumoniae polypeptide (or fragments) of 80 - 680 amino
FT acids in length that comprise at least one histidine triad residue
FT (HxxHxH) or a coiled-coil region, or an antibody directed against
FT these features. The vaccine is useful in protecting against infection
FT by Streptococcus pneumoniae. The vaccine composition comprising
FT antibodies to is useful for passive immunization for treating
FT CC pneumococcal infections which includes otitis media, nasopharyngeal
FT and bronchial infections.
FT
FT Sequence 819 AA;
SQ

```

```

Query Match 87.3%; Score 695; DB 21; Length 819;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 SYELGLYQARTVYKNNRVSYSIDGKATOKTENLTPEDEYSKRKGINAEQIVIKITDGYVT 60
|
|
|
Db 21 SYELGLYQARTVYKNNRVSYSIDGKATOKTENLTPEDEYSKRKGINAEQIVIKITDGYVT 80
|
|
|
QY 61 SHGDHYHYNGKVPYDAIISEELMKDPNYKLEDEIVNEVGGVYIKVDGKYVYLKDA 120
|
|
|
Db 81 SHGDHYHYNGKVPYDAIISEELMKDPNYKLEDEIVNEVGGVYIKVDGKYVYLKDA 140
|
|
|
QY 121 AHADVNRTEEINRQKQESHQREGGTFRNDGAVALARSGRYTTDDGYIFNASDIIEBT 180
|
|
|
Db 141 AHADVNRTEEINRQKQESHQREGGTFRNDGAVALARSGRYTTDDGYIFNASDIIEBT 200
|
|
|
QY 181 GDAYIVPHGDHYHYPKNELSASELAAPFLSGRGNLSNSTRYRQNSDNTSRTNWVS 240
|
|
|

```

```

Db      201 GDAYVPHGDHYHYIPKNELSASELAAABAFISGRGNLSNSTYRRQNSDNTSRTNWVPS 260
Qy      241 VSNQGTNTNTSNNTSNTSOASQSDNDISLKQYLKPLSORHVSDDLVDPAQITRT 300
Db      261 VSNQGTNTNTSNNTSNTSOASQSDNDISLKQYLKPLSORHVSDDLVDPAQITRT 320
Qy      301 ARGVAVPHGDHYHYIPYSQMSLEERIRIIRPLRFRSNHWVPDSRPEQSPQPTPEPSPG 360
Db      321 ARGVAVPHGDHYHYIPYSQMSLEERIRIIRPLRFRSNHWVPDSRPEQSPQPTPEPSPG 380
Qy      361 POPAPNLKIDSNSSLSVQLVRKVGEGYVPEEKGISRYVPAKDLPSSETVKNLESKLSKQES 420
Db      381 POPAPNLKIDSNSSLSVQLVRKVGEGYVPEEKGISRYVPAKDLPSSETVKNLESKLSKQES 440
Qy      421 VSHLTAKENAVAPPDQEFYDQKAVNLLTEAHKALFXNKGKNSDFQALDKLERLNDESTN 480
Db      441 VSHLTAKENAVAPPDQEFYDQKAVNLLTEAHKALFXNKGKNSDFQALDKLERLNDESTN 500
Qy      481 KEKLVDDLALFAPITTPHERLQKPNQSELEYTEDEVRIQLADKYTSDGYTFDEHDIISD 540
Db      501 KEKLVDDLALFAPITTPHERLQKPNQSELEYTEDEVRIQLADKYTSDGYTFDEHDIISD 560
Qy      541 EGDAYVTPHMGSHWIGKDSLSDKEKVAQAAYTEKGLPPSPDADVKAAPTGSAAIY 600
Db      561 EGDAYVTPHMGSHWIGKDSLSDKEKVAQAAYTEKGLPPSPDADVKAAPTGSAAIY 620
Qy      601 NRVAGEKRIPLVRIPYMEVHEVVENGNLIIPIKHCHYNIKFAMWDDHTYKAPNGYTLED 660
Db      621 NRVAGEKRIPLVRIPYMEVHEVVENGNLIIPIKHCHYNIKFAMWDDHTYKAPNGYTLED 680
Qy      661 LFATIKYVVEHPDERPHSNDGKGNASEHVLGKKDSEDPNNKFKADEEPEVEETPAPEVP 720
Db      681 LFATIKYVVEHPDERPHSNDGKGNASEHVLGKKDSEDPNNKFKADEEPEVEETPAPEVP 740
Qy      721 QVETEKVAQKKEAEVLLAKYTDSSLKANATETLAGLRNNLTLOIMDNNSIMAEKELLA 780
Db      741 QVETEKVAQKKEAEVLLAKYTDSSLKANATETLAGLRNNLTLOIMDNNSIMAEKELLA 800
Qy      781 LKGSNPSVSKEKIN 796
Db      801 LKGSNPSVSKEKIN 816

RESULT 5
AAB12728
ID AAB12728 standard; Proteoin, 334 AA.
XX
AC AAB12728;
XX
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae BVH-11A protein antigen SEQ ID NO:61.
XX
KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.
XX
OS Streptococcus pneumoniae.
XX
PN WO200039299-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA01218.
XX
PR 23-DEC-1998; 98US-0113800.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
XX
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX
DR WPI; 2000-452397/39.

```

```

XX      Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT      otitis media, bacteraemia and/or pneumonia -
XX
XX      Claim 18; Fig 26; 106pp; English.
XX
CC      The present invention describes nucleic acids (I) encoding protein
CC      antigens (II) from Streptococcus pneumoniae. The protein antigens
CC      have bactericidal activity. The nucleic acids, encoding the proteins
CC      antigens, may be used for the recombinant production of the proteins
CC      they encode. The protein antigens may then be used as vaccines for the
CC      prevention and treatment of Streptococcal infections in mammals
CC      (especially humans) which result in, e.g. meningitis, otitis media,
CC      bacteraemia and/or pneumonia. The present sequence represents the
CC      S. pneumoniae BVH-11A protein antigen.
XX
SQ      Sequence 334 AA:
XX
Query Match 10.1%; Score 80; DB 21; Length 334;
Best Local Similarity 100.0%; Pred. No. 2,6e-68;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      11 TVKENNRVSYIDGKQATQKTEMLTPDEVSKREGINAEQIVYKITDQGVYSHGHHYNN 70
Db      12 TVKENNRVSYIDGKQATQKTEMLTPDEVSKREGINAEQIVYKITDQGVYSHGHHYNN 71
Qy      71 GKVPYDAIISBELMKDPNY 90
Db      72 GKVPYDAIISBELMKDPNY 91

RESULT 6
AAU84027
ID AAU84027 standard; Peptide; 334 AA.
XX
AC AAU84027;
XX
DT 08-MAY-2002 (first entry)
XX
DE Truncated variant of S. pneumoniae BVH-11, BVH-11A.
XX
KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutein.
XX
OS Streptococcus pneumoniae.
XX
PN WO200198334-A2.
XX
PD 27-DEC-2001.
XX
PF 19-JUN-2001; 2001WO-CA00908.
XX
PR 20-JUN-2000; 2000US-212683P.
XX
PA (SHIR-) SHIRE BIOCHEM INC.
XX
PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX
DR WPI; 2002-122272/16.
XX
DE New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT      epitope-bearing polypeptides, useful as vaccine components for treating
PT      or preventing streptococcal infections such as otitis media,
XX      meningitis, and bacteraemia -
XX
XX      Example 1; Page -; 113pp; English.
XX
CC      The invention describes an isolated polypeptide (I) with 70-90%
CC      identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC      BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC      comprising (I) is useful for therapeutic or prophylactic treatment of
CC      meningitis, otitis media, bacteraemia or pneumonia infection in an

```

individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by Streptococcus pneumoniae, group A Streptococcus such as Streptococcus pyogenes, group B Streptococcus such as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. noecardia or Streptococcus aureus) in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The Streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (III) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating S. pneumonia nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a truncate of a Streptococcus pneumoniae gene used to obtain antigenic peptides, described in the method of the invention.

Note: This sequence does not appear in the specification but has been created according to information given in the invention.

Sequence 334 AA;

Query Match 10.1%; Score 80; DB 23; Length 334;
Best Local Similarity 100.0%; Pred. No. 2.6e-68;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 TVKENNRVSYIDGKATQKTEMLTPDEVSKREGINAEOIVIKITDQGYVTSQGHYHYNN 70
DB 12 TVKENNRVSYIDGKATQKTEMLTPDEVSKREGINAEOIVIKITDQGYVTSQGHYHYNN 71

QY 71 GKVPYDAIISBELMKDPNY 90
DB 72 GKVPYDAIISBELMKDPNY 91

RESULT 7
AAB12745
ID AAB12745 standard; Protein; 690 AA.
AC AAB12745;
XX
XX
DT 23-NOV-2000 (first entry)
XX
XX Streptococcus pneumoniae NEM16 protein antigen SEQ ID NO:79.
DE Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.
XX
OS Streptococcus pneumoniae.
XX
XX WO200039299-A2.
XX
XX 06-JUL-2000.
XX
XX 20-DEC-1999; 99WO-CA01218.
XX
XX 23-DEC-1998; 98US-0113800.
XX
XX (BIOC-) BIOCHEM PHARMA INC.
XX
XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
PI WPI, 2000-452397/39.
XX
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia -
XX
XX Claim 18; Fig 44; 106pp; English.
XX
XX The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the

prevention and treatment of Streptococcal infections in mammals (especially humans) which result in, e.g. meningitis, otitis media, bacteraemia and/or pneumonia. The present sequence represents the S. pneumoniae NEM16 protein antigen.

Sequence 690 AA;

Query Match 10.1%; Score 80; DB 21; Length 690;
Best Local Similarity 100.0%; Pred. No. 5e-68;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 TVKENNRVSYIDGKATQKTEMLTPDEVSKREGINAEOIVIKITDQGYVTSQGHYHYNN 70
DB 12 TVKENNRVSYIDGKATQKTEMLTPDEVSKREGINAEOIVIKITDQGYVTSQGHYHYNN 71

QY 71 GKVPYDAIISBELMKDPNY 90
DB 72 GKVPYDAIISBELMKDPNY 91

RESULT 8
AAU84093
ID AAU84093 standard; Peptide; 690 AA.
AC AAU84093;
XX
XX
DT 08-MAY-2002 (first entry)
XX
XX Truncated variant of S. pneumoniae BVH-11, NEM16.
DE BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; munein.
XX
OS Streptococcus pneumoniae.
XX
XX WO200198334-A2.
XX
XX 27-DEC-2001.
XX
XX 19-JUN-2001; 2001WO-CA00908.
XX
XX 20-JUN-2000; 2000US-212683P.
XX
XX (SHIR-) SHIRE BIOCHEM INC.
XX
XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
PI WPI, 2002-122272/16.
XX
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
XX meningitis, and bacteraemia -
XX
XX Example 1; Page -; 113pp; English.
XX
XX The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. noecardia or
CC Streptococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for

CC diagnosing streptococcal infections. This sequence represents a truncate
 CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
 CC described in the method of the invention.
 CC Note: This sequence does not appear in the specification but has
 CC been created according to information given in the invention.

XX Sequence 690 AA;

Query Match 10.1%; Score 80; DB 23; Length 690;
 Best Local Similarity 100.0%; Pred. No. 5e-68;
 Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 TVKENNRVSYIDGKATOKTENLTPDEVSKREGINAQIVIKITDQGYVTSHGHHYNN 70
 Db 12 TVKENNRVSYIDGKATOKTENLTPDEVSKREGINAQIVIKITDQGYVTSHGHHYNN 71

Qy 71 GKVPYDAIISEELMKDPNY 90
 Db 72 GKVPYDAIISEELMKDPNY 91

RESULT 9

AA12727
 ID AAB12727 standard; Protein; 821 AA.

AC AAB12727;

DT 21-NOV-2000 (first entry)

XX Streptococcus pneumoniae BVH-11M protein antigen SEQ ID NO:60.

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KW otitis media; pneumonia; immunisation; bactericidal.

XX Streptococcus pneumoniae.

PN WO200039299-A2.

PD 06-JUL-2000.

PF 20-DEC-1999; 99WO-CA01218.

PR 23-DEC-1998; 98US-0113800.

PA (BIOC-) BIOCHEM PHARMA INC.

P1 Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

DR WPI; 2000-452397/39.

XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
 FT otitis media, bacteraemia and/or pneumonia -

PS Claim 18; Fig 25; 106pp; English.

XX The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the protein
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence represents the
 CC S. pneumoniae BVH-11M protein antigen.

XX Sequence 821 AA;

Query Match 10.1%; Score 80; DB 21; Length 821;

Best Local Similarity 100.0%; Pred. No. 5.8e-68;
 Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 TVKENNRVSYIDGKATOKTENLTPDEVSKREGINAQIVIKITDQGYVTSHGHHYNN 70

Db 12 TVKENNRVSYIDGKATOKTENLTPDEVSKREGINAQIVIKITDQGYVTSHGHHYNN 71
 Qy 71 GKVPYDAIISEELMKDPNY 90
 Db 72 GKVPYDAIISEELMKDPNY 91

RESULT 10

AA12766
 ID AAB12766 standard; Protein; 821 AA.

AC AAB12766;

DT 21-NOV-2000 (first entry)

XX Streptococcus pneumoniae strain SP64 BVH-11 protein antigen.

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KW otitis media; pneumonia; immunisation; bactericidal.

XX Streptococcus pneumoniae.

PN WO200039299-A2.

PD 06-JUL-2000.

PF 20-DEC-1999; 99WO-CA01218.

PR 23-DEC-1998; 98US-0113800.

PA (BIOC-) BIOCHEM PHARMA INC.

P1 Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

DR WPI; 2000-452397/39.

XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
 FT otitis media, bacteraemia and/or pneumonia -

PS Disclosure; Fig 12; 106pp; English.

XX The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the protein
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence represents a
 CC S. pneumoniae BVH-11 protein antigen, from the present invention.

XX Sequence 821 AA;

Query Match 10.1%; Score 80; DB 21; Length 821;
 Best Local Similarity 100.0%; Pred. No. 5.8e-68;
 Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 TVKENNRVSYIDGKATOKTENLTPDEVSKREGINAQIVIKITDQGYVTSHGHHYNN 70
 Db 12 TVKENNRVSYIDGKATOKTENLTPDEVSKREGINAQIVIKITDQGYVTSHGHHYNN 71

Qy 71 GKVPYDAIISEELMKDPNY 90
 Db 72 GKVPYDAIISEELMKDPNY 91

RESULT 11

AAU84026
 ID AAU84026 standard; Peptide; 821 AA.

AC AAU84026;

08-MAY-2002 (first entry)

Truncated variant of *S. pneumoniae* BVH-11, BVH-11M.

BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia; pneumonia; streptococcal bacterial infection; mutant; mutcin.

Streptococcus pneumoniae.

Synthetic.

MO200198334-A2.

27-DEC-2001.

19-JUN-2001; 2001WO-CA00908.

20-JUN-2000; 2000US-212683P.

(SHR-) SHRE BIOCHEM INC.

Hamel J, Ouellet C, Charland N, Martin D, Brodeur B; WPI, 2002-122272/16.

New *Streptococcus pneumoniae* BVH-3 and BVH-11 variant and epitope-bearing polypeptides, useful as vaccine components for treating or preventing streptococcal infections such as otitis media, meningitis, and bacteraemia

Example 1; Page -, 113pp: English.

The invention describes an isolated polypeptide (I) with 70-90% identity to *Streptococcus pneumoniae* protein BVH-3, BVH-11, variants of BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II) comprising (i) is useful for therapeutic or prophylactic treatment of meningitis, otitis media, bacteraemia or pneumonia infection in an individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by *Streptococcus pneumoniae*, group A streptococcus such as *Streptococcus pyogenes*, group B streptococcus such as *Streptococcus agalactiae*, *S. dysgalactiae*, *S. uberis*, *S. novarcia* or *Streptococcus aureus*) in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The *Streptococcus* polypeptides are useful in a diagnostic test for *S. pneumoniae* infection. (III) is useful for designing DNA probes for use in detecting the presence of *Streptococcus* in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating *S. pneumoniae* nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a truncate of a *Streptococcus pneumoniae* gene used to obtain antigenic peptides, described in the method of the invention.

Note: This sequence does not appear in the specification but has been created according to information given in the invention.

Sequence 821 AA:

Query Match 10.1%; Score 80; DB 23; Length 821;
Best Local Similarity 100.0%; Pred. No. 5.8e-68;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0

11 TVKENNRVSYIDGKQATOKTENLTLPDESVKREGINAEQIVIKITDQGVYTSQDHYHYNN 70
12 TVKENNRVSYIDGKQATOKTENLTLPDESVKREGINAEQIVIKITDQGVYTSQDHYHYNN 71

71 GKVPYDAIISSEELMKDPMY 90
72 GKVPYDAIISSEELMKDPMY 91

RESULT 12
AAB12716
ID AAB12716 standard; Protein; 840 AA.

XX	AA012716;	
AC		
XX		
DT	21-NOV-2000	(first entry)
XX		
XX		
DE	Streptococcus pneumoniae BVH-11 protein antigen SEQ ID NO:4.	
XX		
KM	Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;	
KW	prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;	
XX	otitis media; pneumonia; immunisation; bactericidal.	
OS	Streptococcus pneumoniae.	
XX		
FN	WO200039299-A2.	
XX		
PD	06-JUL-2000.	
XX		
PF	20-DEC-1999; 99WO-CA01218.	
XX		
PR	23-DEC-1998; 98US-0113800.	
XX		
PA	(BIOC-) BIOCHEM PHARMA INC.	
XX		
PI	Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;	
XX	WPI; 2000-452397/39.	
DR	N-PSDB; AAA65731.	
PT	Streptococcal antigens useful for vaccinating against e.g. meningitis,	
XX	otitis media, bacteraemia and/or pneumonia -	
PS	Claim 18; Fig 4; 106pp; English.	
XX		
CC	The present invention describes nucleic acids (I) encoding protein	
CC	antigens (II) from Streptococcus pneumoniae. The protein antigens	
CC	have bactericidal activity. The nucleic acids, encoding the protein	
CC	antigens, may be used for the recombinant production of the proteins	
CC	they encode. The protein antigens may then be used as vaccines for the	
CC	prevention and treatment of Streptococcal infections in mammals	
CC	(especially humans) which result in, e.g. meningitis, otitis media,	
CC	bacteraemia and/or pneumonia. The present sequence represents the	
CC	S. pneumoniae BVH-11 protein antigen.	
XX		
SQ	Sequence 840 AA;	
	Query Match 10.1%; Score 80; DB 21; Length 840;	
	Best Local Similarity 100.0%; Pred. No. 6e-68;	
	Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
QY	11 TVKENNVSVYIDGKQATOKTENLTPDEYSKKEGINAEQIVIKITDQGYVTSBGHYYHN 70	
DB	31 TVKENNVSVYIDGKQATOKTENLTPDEYSKKEGINAEQIVIKITDQGYVTSBGHYYHN 90	
QY	71 GKVPYDAIISEELMKDPNY 90	
DB	91 GKVPYDAIISEELMKDPNY 110	
	RESULT 13	
XX	AAU75933	
XX	ID AAU75933 standard; Protein; 840 AA.	
XX	AAU75933;	
XX		
DT	08-MAY-2002 (first entry)	
XX		
DE	Streptococcus pneumoniae BVH-11 protein.	
XX		
KM	BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;	
KW	pneumonia; streptococcal bacterial infection.	
XX		
OS	Streptococcus pneumoniae.	
XX		

XX Disclosure; Fig 12; 106pp; English.
PS
XX The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents a
CC S. pneumoniae BVH-11 protein antigen, from the present invention.
XX
SQ Sequence 807 AA;
Query Match 7.5%; Score 60; DB 21; Length 807;
Best Local Similarity 100.0%; Pred. No. 1.2e-48;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 31 ENLTPDEVSKREGINAEOIVIKITDQGYVTSBGHGHYNGKVPYDAIISEBLMKDPNY 90
DB 33 ENLTPDEVSKREGINAEOIVIKITDQGYVTSBGHGHYNGKVPYDAIISEBLMKDPNY 92
RESULT 16
AAB12760
ID AAB12760 standard; Protein; 811 AA.
XX
AC AAB12760;
XX
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae strain P4241 BVH-11 protein antigen.
XX
KM Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KM prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KM otitis media; pneumonia; immunisation; bactericidal.
XX
OS Streptococcus pneumoniae.
XX
PM WO200039299-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA01218.
XX
PR 23-DEC-1998; 98US-0113800.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX
DR WPI; 2000-452397/39.
XX
PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia -
XX
PS Disclosure; Fig 12; 106pp; English.
XX
XX The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents a
CC S. pneumoniae BVH-11 protein antigen, from the present invention.
XX
SQ Sequence 811 AA;
Query Match 7.5%; Score 60; DB 21; Length 811;
Best Local Similarity 100.0%; Pred. No. 1.2e-48;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 31 ENLTPDEVSKREGINAEOIVIKITDQGYVTSBGHGHYNGKVPYDAIISEBLMKDPNY 90
DB 33 ENLTPDEVSKREGINAEOIVIKITDQGYVTSBGHGHYNGKVPYDAIISEBLMKDPNY 92
RESULT 18
AAB12762
ID AAB12762 standard; Protein; 811 AA.
XX
AC AAB12762;
XX
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae strain A66 BVH-11 protein antigen.

Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 31 ENLTPDEVSKREGINAEOIVIKITDQGYVTSBGHGHYNGKVPYDAIISEBLMKDPNY 90
DB 33 ENLTPDEVSKREGINAEOIVIKITDQGYVTSBGHGHYNGKVPYDAIISEBLMKDPNY 92
RESULT 17
AAB12761
ID AAB12761 standard; Protein; 811 AA.
XX
AC AAB12761;
XX
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae strain WU2 BVH-11 protein antigen.
XX
KM Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KM prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KM otitis media; pneumonia; immunisation; bactericidal.
XX
OS Streptococcus pneumoniae.
XX
PM WO200039299-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA01218.
XX
PR 23-DEC-1998; 98US-0113800.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX
DR WPI; 2000-452397/39.
XX
PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia -
XX
PS Disclosure; Fig 12; 106pp; English.
XX
XX The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents a
CC S. pneumoniae BVH-11 protein antigen, from the present invention.
XX
SQ Sequence 811 AA;
Query Match 7.5%; Score 60; DB 21; Length 811;
Best Local Similarity 100.0%; Pred. No. 1.2e-48;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 31 ENLTPDEVSKREGINAEOIVIKITDQGYVTSBGHGHYNGKVPYDAIISEBLMKDPNY 90
DB 33 ENLTPDEVSKREGINAEOIVIKITDQGYVTSBGHGHYNGKVPYDAIISEBLMKDPNY 92
RESULT 18
AAB12762
ID AAB12762 standard; Protein; 811 AA.
XX
AC AAB12762;
XX
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae strain A66 BVH-11 protein antigen.

KM Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KM prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KM otitis media; pneumonia; immunisation; bactericidal.
XX
OS Streptococcus pneumoniae.
XX
PN WO200039299-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA01218.
XX
PR 23-DEC-1998; 98US-0113800.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
XX
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX
DR WPI; 2000-452397/39.
XX
PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia -
XX
PS Disclosure; Fig 12; 106pp; English.
XX
CC The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents a
CC S. pneumoniae BVH-11 protein antigen, from the present invention.
XX
SQ Sequence 811 AA;
XX
Query Match 7.5%; Score 60; DB 21; Length 811;
Best Local Similarity 100.0%; Pred. No. 1.2e-48;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYNGKVPYDAIISEELMKDPNY 90
33 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYNGKVPYDAIISEELMKDPNY 92
Db
RESULT 19
AAB12763
ID AAB12763 standard; Protein; 811 AA.
XX
AC AAB12763;
XX
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae strain RX1 BVH-11 protein antigen.
XX
KM Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KM prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KM otitis media; pneumonia; immunisation; bactericidal.
XX
OS Streptococcus pneumoniae.
XX
PN WO200039299-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA01218.
XX
PR 23-DEC-1998; 98US-0113800.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

XX
DR WPI; 2000-452397/39.
XX
PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia -
XX
PS Disclosure; Fig 12; 106pp; English.
XX
CC The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the proteins
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents a
CC S. pneumoniae BVH-11 protein antigen, from the present invention.
XX
SQ Sequence 811 AA;
XX
Query Match 7.5%; Score 60; DB 21; Length 811;
Best Local Similarity 100.0%; Pred. No. 1.2e-48;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYNGKVPYDAIISEELMKDPNY 90
33 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYNGKVPYDAIISEELMKDPNY 92
Db
RESULT 20
AAB12756
ID AAB12756 standard; Protein; 816 AA.
XX
AC AAB12756;
XX
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae strain P4241 BVH-11-2 protein antigen.
XX
KM Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KM prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KM otitis media; pneumonia; immunisation; bactericidal.
XX
OS Streptococcus pneumoniae.
XX
PN WO200039299-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA01218.
XX
PR 23-DEC-1998; 98US-0113800.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX
DR WPI; 2000-452397/39.
XX
PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia -
XX
PS Disclosure; Fig 12; 106pp; English.
XX
CC The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the proteins
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents a
CC S. pneumoniae BVH-11-2 protein antigen, from the present invention.

```

XX SQ      Sequence      816 AA;
XX
XX Query Match      7.5%; Score 60; DB 21; Length 816;
XX Best Local Similarity 100.0%; Pred. No. 1.2e-48;
XX Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY      31 ENLTPDEVSKREGINAEOQIVIKITDQGYVTSHGHDHYHYNGKVPYDAIISBELMKDPNY 90
XX      |||
XX      33 ENLTPDEVSKREGINAEOQIVIKITDQGYVTSHGHDHYHYNGKVPYDAIISBELMKDPNY 92
XX
Db
XX
XX RESULT 21
XX AAB12757
XX ID      AAB12757 standard; Protein; 816 AA.
XX
XX AC      AAB12757;
XX
XX DT      21-NOV-2000 (first entry)
XX
XX DE      Streptococcus pneumoniae strain A66 BVH-11-2 protein antigen.
XX
XX KM      Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
XX prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX otitis media; pneumonia; immunisation; bactericidal.
XX
XX OS      Streptococcus pneumoniae.
XX
XX PN      WO200039299-A2.
XX
XX PD      06-JUL-2000.
XX
XX PF      20-DEC-1999; 99WO-CA01218.
XX
XX PR      23-DEC-1998; 98US-0113800.
XX
XX PA      (BIOC-) BIOCHEM PHARMA INC.
XX
XX PI      Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX
XX DR      WPI; 2000-452397/39.
XX
XX PT      Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX otitis media, bacteraemia and/or pneumonia -
XX
XX PS      Disclosure; Fig 12; 106pp; English.
XX
XX CC      The present invention describes nucleic acids (I) encoding protein
XX antigens (II) from Streptococcus pneumoniae. The protein antigens
XX have bactericidal activity. The nucleic acids, encoding the protein
XX antigens, may be used for the recombinant production of the proteins
XX they encode. The protein antigens may then be used as vaccines for the
XX prevention and treatment of Streptococcal infections in mammals
XX (especially humans) which result in, e.g. meningitis, otitis media,
XX bacteraemia and/or pneumonia. The present sequence represents a
XX S. pneumoniae BVH-11-2 protein antigen, from the present invention.
XX
XX SQ      Sequence      816 AA;
XX
XX Query Match      7.5%; Score 60; DB 21; Length 816;
XX Best Local Similarity 100.0%; Pred. No. 1.2e-48;
XX Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY      31 ENLTPDEVSKREGINAEOQIVIKITDQGYVTSHGHDHYHYNGKVPYDAIISBELMKDPNY 90
XX      |||
XX      33 ENLTPDEVSKREGINAEOQIVIKITDQGYVTSHGHDHYHYNGKVPYDAIISBELMKDPNY 92
XX
Db
XX
XX RESULT 22
XX AAB12758
XX ID      AAB12758 standard; Protein; 816 AA.
XX
XX AC      AAB12758;
XX

```

```

XX XX      21-NOV-2000 (first entry)
XX DT
XX DE      Streptococcus pneumoniae strain WU2 BVH-11-2 protein antigen.
XX
XX KM      Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
XX prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX otitis media; pneumonia; immunisation; bactericidal.
XX
XX OS      Streptococcus pneumoniae.
XX
XX PN      WO200039299-A2.
XX
XX PD      06-JUL-2000.
XX
XX PF      20-DEC-1999; 99WO-CA01218.
XX
XX PR      23-DEC-1998; 98US-0113800.
XX
XX PA      (BIOC-) BIOCHEM PHARMA INC.
XX
XX PI      Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX
XX DR      WPI; 2000-452397/39.
XX
XX PT      Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX otitis media, bacteraemia and/or pneumonia -
XX
XX PS      Disclosure; Fig 12; 106pp; English.
XX
XX CC      The present invention describes nucleic acids (I) encoding protein
XX antigens (II) from Streptococcus pneumoniae. The protein antigens
XX have bactericidal activity. The nucleic acids, encoding the protein
XX antigens, may be used for the recombinant production of the proteins
XX they encode. The protein antigens may then be used as vaccines for the
XX prevention and treatment of Streptococcal infections in mammals
XX (especially humans) which result in, e.g. meningitis, otitis media,
XX bacteraemia and/or pneumonia. The present sequence represents a
XX S. pneumoniae BVH-11-2 protein antigen, from the present invention.
XX
XX SQ      Sequence      816 AA;
XX
XX Query Match      7.5%; Score 60; DB 21; Length 816;
XX Best Local Similarity 100.0%; Pred. No. 1.2e-48;
XX Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY      31 ENLTPDEVSKREGINAEOQIVIKITDQGYVTSHGHDHYHYNGKVPYDAIISBELMKDPNY 90
XX      |||
XX      33 ENLTPDEVSKREGINAEOQIVIKITDQGYVTSHGHDHYHYNGKVPYDAIISBELMKDPNY 92
XX
Db
XX
XX RESULT 23
XX AAB12740
XX ID      AAB12740 standard; Protein; 819 AA.
XX
XX AC      AAB12740;
XX
XX DT      21-NOV-2000 (first entry)
XX
XX DE      Streptococcus pneumoniae BVH-11-2M protein antigen SEQ ID NO:73.
XX
XX KM      Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
XX prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX otitis media; pneumonia; immunisation; bactericidal.
XX
XX OS      Streptococcus pneumoniae.
XX
XX PN      WO200039299-A2.
XX
XX PD      06-JUL-2000.
XX
XX PF      20-DEC-1999; 99WO-CA01218.
XX

```

PR 23-DEC-1998; 98US-0113800.
 XX (BIOC-) BIOCHEM PHARMA INC.
 PA Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 PI WPI; 2000-452397/39.
 XX
 DR Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteraemia and/or pneumonia -
 XX
 PS Claim 18; Fig 38; 106pp; English.
 XX
 CC The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the protein
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence represents the
 CC S. pneumoniae BVH-11-2M protein antigen.
 XX
 SQ Sequence 819 AA;
 XX
 Query Match 7.5%; Score 60; DB 21; Length 819;
 Best Local Similarity 100.0%; Pred. No. 1.2e-48;
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 31 ENLTPDEVSKREGINAEQIVKITDQGYVTSBGHGHYHYNGKVPYDAIISELLMKDPNY 90
 Db 33 ENLTPDEVSKREGINAEQIVKITDQGYVTSBGHGHYHYNGKVPYDAIISELLMKDPNY 92
 XX
 RESULT 24
 AAB12754
 ID AAB12754 standard; Protein; 819 AA.
 XX
 AC AAB12754;
 XX
 DT 21-NOV-2000 (first entry)
 XX
 DE Streptococcus pneumoniae strain SP64 BVH-11-2 protein antigen.
 XX
 KM Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen, vaccine;
 KM prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KM otitis media; pneumonia; immunisation; bactericidal.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200039299-A2.
 XX
 PD 06-JUL-2000.
 XX
 PF 20-DEC-1999; 99WO-CA01218.
 XX
 PR 23-DEC-1998; 98US-0113800.
 XX
 PA (BIOC-) BIOCHEM PHARMA INC.
 XX
 PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 DR WPI; 2000-452397/39.
 XX
 PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteraemia and/or pneumonia -
 XX
 PS Disclosure; Fig 12; 106pp; English.
 XX
 CC The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the protein
 CC antigens, may be used for the recombinant production of the proteins

CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence represents a
 CC S. pneumoniae BVH-11-2 protein antigen, from the present invention.
 XX
 SQ Sequence 819 AA;
 XX
 Query Match 7.5%; Score 60; DB 21; Length 819;
 Best Local Similarity 100.0%; Pred. No. 1.2e-48;
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 31 ENLTPDEVSKREGINAEQIVKITDQGYVTSBGHGHYHYNGKVPYDAIISELLMKDPNY 90
 Db 33 ENLTPDEVSKREGINAEQIVKITDQGYVTSBGHGHYHYNGKVPYDAIISELLMKDPNY 92
 XX
 RESULT 25
 AAB01469
 ID AAB01469 standard; Protein; 819 AA.
 XX
 AC AAB01469;
 XX
 DT 20-OCT-2000 (first entry)
 XX
 DE Recombinant variant of Sp36 (Sp36B) of S. pneumoniae.
 XX
 KM Streptococcus pneumoniae; infection; vaccine; coiled coil region;
 KM histidine triad residue; Sp36; antibody; otitis media;
 KM nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
 KM meningitis; lobar pneumonia.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200037105-A2.
 XX
 PD 29-JUN-2000.
 XX
 PF 21-DEC-1999; 99WO-US30390.
 XX
 PR 21-DEC-1998; 98US-0113048.
 XX
 PA (MEDI-) MEDIMUNE INC.
 XX
 PI Johnson LS, Koenig S, Adamou JB;
 DR WPI; 2000-452129/39.
 DR N-PSDB; AAA47605.
 XX
 PT Vaccine useful for prophylaxis and treatment of pneumococcal infections
 PT such as otitis media, nasopharyngeal and bronchial infections,
 PT comprises Streptococcus pneumoniae proteins
 XX
 PS Claim 1; Page 65-69; 70pp; English.
 XX
 CC Although a number of proteins have been suggested as being involved
 CC in the pathogenicity of Streptococcus pneumoniae, there still remains
 CC a need to identify polypeptides having epitopes in common from
 CC various strains of S. pneumoniae in order to utilise such
 CC polypeptides in vaccines to protect against a wide variety of
 CC S. pneumoniae. New vaccine compositions are described which comprise a
 CC Streptococcus pneumoniae polypeptide (or fragments) of 80 - 680 amino
 CC acids in length that comprise at least one histidine triad residue
 CC (HxxHxx) or a coiled-coil region, or an antibody directed against
 CC these features. The vaccine is useful in protecting against infection
 CC by Streptococcus pneumoniae. The vaccine composition comprising
 CC antibodies to is useful for passive immunization for treating
 CC pneumococcal infections which includes otitis media, nasopharyngeal
 CC and bronchial infections.
 XX
 SQ Sequence 819 AA;
 XX
 Query Match 7.5%; Score 60; DB 21; Length 819;

Best Local Similarity 100.0%; Pred. No. 1.2e-48;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 ENLTPEVSKREGINAEOIVIKITDQGYVTSFGDHYHYNGVPPYDAITSEELMKDPNY 90
Db 52 ENLTPEVSKREGINAEOIVIKITDQGYVTSFGDHYHYNGVPPYDAITSEELMKDPNY 111

RESULT 26
AAU84087

ID AAU84087 standard; Peptide; 819 AA.

XX AAU84087;

DT 08-MAY-2002 (first entry)

DE Truncated variant of S. pneumoniae BVH-11-2, BVH-

XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;

KW pneumonia; streptococcal bacterial infection; mutant; mutein.

OS Streptococcus pneumoniae.

PN WO200198334-A2.

PD 27-DEC-2001.

PF 19-JUN-2001; 2001WO-CA00908.

PR 20-JUN-2000; 2000US-212683P.

PA (SHIR-) SHIRE BIOCHEM INC.

PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

DR WPI; 2002-122272/16.

PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia -

PS Example 1; Page -; 113pp; English.

XX The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novacardia or
CC Streptococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a truncate
CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
CC described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.

XX Sequence 819 AA;

Query Match 7.5%; Score 60; DB 23; Length 819;
Best Local Similarity 100.0%; Pred. No. 1.2e-48;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 ENLTPEVSKREGINAEOIVIKITDQGYVTSFGDHYHYNGVPPYDAITSEELMKDPNY 90
Db 33 ENLTPEVSKREGINAEOIVIKITDQGYVTSFGDHYHYNGVPPYDAITSEELMKDPNY 92

RESULT 27
ABU01597

ID ABU01597 standard; Protein; 819 AA.

XX ABU01597;

DT 11-FEB-2003 (first entry)

DE S. pneumoniae type 4 strain protein from coding region #1173.

KW Bacterial meningitis; pneumonia; sepsis; otitis media;

KW ear infection; antiinflammatory; antibacterial; immunostimulant;

XX auditory; respiratory; gene therapy; vaccine.

OS Streptococcus pneumoniae type 4 strain.

PN WO200277021-A2.

PD 03-OCT-2002.

PF 27-MAR-2002; 2002WO-IB02163.

PR 27-MAR-2001; 2001GB-0007658.

PA (CHIR-) CHIRON SPA.

PI Masignani V, Tettelin H, Frazer C;

DR WPI; 2003-040579/03.

PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT useful as medicaments for treating or preventing a disease or infection
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
PT or ear infection -

PS Claim 1; SEQ ID No 2346; 56pp; English.

XX The invention relates to a protein comprising or having at least 50%
CC identity to any of the 2469 amino acid sequences, identified in the
CC specification (available on a computer readable format), or its fragment,
CC expressed from 2469 of 2489 identified DNA coding regions from the
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC ABS6454. Also included are an antibody which binds one of the
CC proteins, treating a patient by administering the protein, DNA or
CC antibody (in a composition), a kit comprising first and second primers,
CC which are the nucleic acid cited above or fragments between nucleotides
CC 8-100 of a sequence not defined in the specification, for amplifying a
CC target sequence contained within a Streptococcus nucleic acid sequence,
CC where the first primer is substantially complementary to the target
CC sequence and the second primer is substantially complementary to the
CC complement of the target sequence, and where the parts of the primers
CC having substantial complementarity define the termini of the target
CC sequence to be amplified, assay comprising contacting a test compound
CC with the protein, and determining whether the test compound binds to the
CC protein and a Streptococcus pneumoniae bacterium, where one or more
CC genes encoding the proteins has been rendered inactive. The proteins,
CC nucleic acid molecules, antibody and compositions are useful as
CC medicaments for treating or preventing a disease or infection due to
CC streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,
CC sepsis, otitis media or ear infection. They are also useful in developing
CC vaccines, diagnostics and antibiotics. The methods are useful for
CC identifying immunodominant proteins. The present sequence is one of
CC the 2469 proteins expressed by the identified coding regions from the
CC genomic sequence.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic

CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX

SO Sequence 819 AA;

Query Match 7.5%; Score 60; DB 24; Length 819;
Best Local Similarity 100.0%; Pred. No. 1.2e-48;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHHYNGKVPYDAIISSELMKDPNY 90
DB 52 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHHYNGKVPYDAIISSELMKDPNY 111

RESULT 28

AAB12755
ID AAB12755 standard; Protein; 820 AA.

AC AAB12755;

DT 21-NOV-2000 (first entry)

DE Streptococcus pneumoniae strain JNR7/87 BVH-11-2 protein antigen.

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;

KM otitis media; pneumonia; immunisation; bactericidal.

XX Streptococcus pneumoniae.

XX WO200039299-A2.

PD 06-JUL-2000.

XX 20-DEC-1999; 99WO-CA01218.

XX 23-DEC-1998; 98US-0113800.

PA (BIOC-) BIOCHEM PHARMA INC.

PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

XX WPI; 2000-452397/39.

PT Streptococcal antigens useful for vaccinating against e.g. meningitis,

XX otitis media, bacteraemia and/or pneumonia.

PS Disclosure; Fig 12; 106pp; English.

XX The present invention describes nucleic acids (I) encoding protein

CC antigens (II) from Streptococcus pneumoniae. The protein antigens

CC have bactericidal activity. The nucleic acids, encoding the protein

CC antigens, may be used for the recombinant production of the proteins

CC they encode. The protein antigens may then be used as vaccines for the

CC prevention and treatment of Streptococcal infections in mammals

CC (especially humans) which result in, e.g. meningitis, otitis media,

CC bacteraemia and/or pneumonia. The present sequence represents a

CC S. pneumoniae BVH-11-2 protein antigen, from the present invention.

XX Sequence 820 AA;

Query Match 7.5%; Score 60; DB 21; Length 820;
Best Local Similarity 100.0%; Pred. No. 1.2e-48;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHHYNGKVPYDAIISSELMKDPNY 90
DB 33 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHHYNGKVPYDAIISSELMKDPNY 92

RESULT 29
AAY91939
ID AAY91939 standard; Protein; 826 AA.

XX AAY91939;

AC AAY91939 (first entry)

DT 19-JUL-2000

XX S. pneumoniae 92 kDa human C3-degrading protein.

DE Human C3-degrading protein; 92 kDa; immunostimulatory; vaccine;

XX inhibitor; inflammation; organ rejection; xenotransplantation.

KW Streptococcus pneumoniae.

XX WO200017370-A1.

XX 30-MAR-2000.

XX 24-SEP-1999; 99WO-US22362.

XX 24-SEP-1998; 98US-0101736.

XX 31-MAR-1999; 99US-0283094.

XX (MNU) UNIV MINNESOTA.

PA (AMCY) AMERICAN CYANAMID CO.

PI Hostetter MK, Finkel DJ, Cheng Q, Green BA, Masi AW;

XX WPI; 2000-283594/24.

XX N-PSDB; AAA08557.

DR Isolated polypeptide is used to stimulate immune system and immunize or

PT treat a mammalian subject against Streptococcus pneumoniae infection or

XX colonization

XX Claim 8; Page 55-57; 63pp; English.

XX The present sequence, isolated from Streptococcus pneumoniae, is a

CC human C3-degrading protein of about 92 kDa. This sequence may encompass

CC a smaller, approximately 20 kDa protein (see AAY91938), also having

CC human C3-degrading activity. The DNA sequences (AAA08556-57) can be

CC used for producing an immune response to Streptococcus pneumoniae in a

CC mammal. Antibodies against the proteins can be used to inhibit

CC S. pneumoniae-mediated C3 degradation. C3-mediated inflammation and

CC rejection in xenotransplantation can be inhibited by expressing the

CC nucleic acid sequences on the surface of an organ of an animal. In

CC particular, the polypeptides are useful for stimulating the immune

CC system and are effective to immunize or treat a mammalian subject

CC against Streptococcus pneumoniae infection or colonization.

XX Sequence 826 AA;

Query Match 7.5%; Score 60; DB 21; Length 826;
Best Local Similarity 100.0%; Pred. No. 1.2e-48;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHHYNGKVPYDAIISSELMKDPNY 90
DB 52 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHHYNGKVPYDAIISSELMKDPNY 111

RESULT 30
AAY81662
ID AAY81662 standard; Protein; 827 AA.

AC AAY81662;

DT 24-MAY-2000 (first entry)

DE Streptococcus pneumoniae protein sequence ID311.

XX Streptococcus pneumoniae; vaccine; screening; protein antigen;

KW antibacterial; antiinflammatory; meningitis; infection; diagnosis;

XX pneumococcal disease.

OS Streptococcus pneumoniae.
 XX
 PN WO200006737-A2.
 XX
 PD 10-FEB-2000.
 XX
 PF 27-JUL-1999; 99WO-GB02451.
 XX
 PR 27-JUL-1998; 98GB-0016337.
 XX
 PR 19-MAR-1999; 99US-0125164.
 XX
 PA (MICR-) MICROBIAL TECHNIQS LTD.
 XX
 PI Gilbert CFG, Hansbro PM;
 XX
 DR WPI; 2000-195300/17.
 XX
 PT New Streptococcal protein, useful as a vaccine, for diagnosis of
 PT pneumococcal diseases and for screening agents capable of antagonizing
 PT or inhibiting expression of the protein -
 XX
 PS Claim 2; Page 99-100; 108pp; English.
 XX
 CC AAY81501 to AAY81679 represent specifically claimed protein sequences
 CC isolated from Streptococcus pneumoniae. AAA05407 to AAA05590 represent
 CC specifically claimed nucleotide sequences isolated from S. pneumoniae.
 CC The sequences have antibacterial and antiinflammatory properties.
 CC The protein sequences, and fragments of them, are useful as immunogens
 CC and/or antigens. The nucleotide sequences can be used in vaccines and in
 CC diagnostic assays. The proteins and nucleotides can be useful for the
 CC detection and diagnosis of S. pneumoniae. The protein sequences are also
 CC useful for screening an agent capable of antagonising, inhibiting or
 CC interfering with the function or expression of the proteins in which the
 CC agent is useful for treatment or prophylaxis of S. pneumoniae infection
 CC and meningitis. AAA05591 to AAA05614 represent primers used in the
 CC exemplification of the present invention.
 CC
 SQ Sequence 827 AA;
 Query Match 7.5%; Score 60; DB 21; Length 827;
 Best Local Similarity 100.0%; Pred. No. 1.2e-48;
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 31 ENLTPDEVSKREGINAEOIVKITDQGYVTSBGDHYHYNGKVPYDAIISELLMKDPNY 90
 |||||
 DB 52 ENLTPDEVSKREGINAEOIVKITDQGYVTSBGDHYHYNGKVPYDAIISELLMKDPNY 111
 |||||
 RESULT 31
 AAB12759
 ID AAB12759 standard; Protein; 834 AA.
 XX
 AC AAB12759;
 XX
 DT 21-NOV-2000 (first entry)
 XX
 DE Streptococcus pneumoniae strain RX1 BVH-11-2 protein antigen.
 XX
 KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KW otitis media; pneumonia; immunisation; bactericidal.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200039299-A2.
 XX
 PD 06-JUL-2000.
 XX
 PF 20-DEC-1999; 99WO-CA01218.
 XX
 PR 23-DEC-1998; 98US-0113800.
 XX
 PA (BIOC-) BIOCHEM PHARMA INC.

XX
 PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 XX
 DR WPI; 2000-452397/39.
 XX
 PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteraemia and/or pneumonia -
 XX
 PS Disclosure; Fig 12; 106pp; English.
 XX
 CC The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the protein
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence represents a
 CC S. pneumoniae BVH-11-2 protein antigen, from the present invention.
 CC
 SQ Sequence 834 AA;
 Query Match 7.5%; Score 60; DB 21; Length 834;
 Best Local Similarity 100.0%; Pred. No. 1.2e-48;
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 31 ENLTPDEVSKREGINAEOIVKITDQGYVTSBGDHYHYNGKVPYDAIISELLMKDPNY 90
 |||||
 DB 33 ENLTPDEVSKREGINAEOIVKITDQGYVTSBGDHYHYNGKVPYDAIISELLMKDPNY 92
 |||||
 RESULT 32
 AAB12720
 ID AAB12720 standard; Protein; 838 AA.
 XX
 AC AAB12720;
 XX
 DT 21-NOV-2000 (first entry)
 XX
 DE Streptococcus pneumoniae BVH-11-2 protein antigen SEQ ID NO:14.
 XX
 KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KW otitis media; pneumonia; immunisation; bactericidal.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200039299-A2.
 XX
 PD 06-JUL-2000.
 XX
 PF 20-DEC-1999; 99WO-CA01218.
 XX
 PR 23-DEC-1998; 98US-0113800.
 XX
 PA (BIOC-) BIOCHEM PHARMA INC.
 XX
 PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 XX
 DR WPI; 2000-452397/39.
 XX
 DR N-PSDB; AAA65737.
 XX
 PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteraemia and/or pneumonia -
 XX
 PS Claim 18; Fig 17; 106pp; English.
 XX
 CC The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the proteins
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals

CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteremia and/or pneumonia. The present sequence represents the
CC S. pneumoniae BVH-11-2 protein antigen.

XX Sequence 838 AA;

Query Match 7.5%; Score 60; DB 21; Length 838;
Best Local Similarity 100.0%; Pred. No. 1.2e-48;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHHYHYNGKVPYDAIISEELMKDPNY 90
Db 52 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHHYHYNGKVPYDAIISEELMKDPNY 111

RESULT 33

AAB01466
ID AAB01466 standard; Protein; 838 AA.

XX AAB01466;

XX 20-OCT-2000 (first entry)

XX Recombinant variant of Sp36 (Sp36D) of S. pneumoniae.

XX Streptococcus pneumoniae; infection; vaccine; coiled coil region;

XX histidine triad residue; Sp36; antibody; otitis media;

XX nasopharyngeal infection; bronchial infection; bronchitis; sepsis;

XX meningitis; lobar pneumonia.

XX Streptococcus pneumoniae.

XX Key location/Qualifiers

FT Region 64..69 /label= Histidine triad residue

FT Region 188..193 /label= Histidine triad residue

FT Region 296..301 /label= Histidine triad residue

FT Region 541..546 /label= Histidine triad residue

FT Region 625..630 /label= Histidine triad residue

XX WO200037105-A2.

XX 29-JUN-2000.

XX 21-DEC-1999; 99WO-US30390.

XX 21-DEC-1998; 98US-0113048.

XX (MEDI-) MEDIMUNE INC.

XX Johnson LS, Koenig S, Adamou JE;

XX MPI: 2000-452129/39.

XX N-PSDB; AAA47602.

XX Vaccine useful for prophylaxis and treatment of pneumococcal infections

XX such as otitis media, nasopharyngeal and bronchial infections,

XX comprises Streptococcus pneumoniae proteins

XX Claim 1; Page 54-57; 70pp; English.

XX Although a number of proteins have been suggested as being involved
XX in the pathogenicity of Streptococcus pneumoniae, there still remains
XX a need to identify polypeptides having epitopes in common from
XX various strains of S. pneumoniae in order to utilise such
XX polypeptides in vaccines to protect against a wide variety of
XX S. pneumoniae. New vaccine compositions are described which comprise a
XX Streptococcus pneumoniae polypeptide (or fragments) of 80 - 680 amino
XX acids in length that comprise at least one histidine triad residue

CC (HxxHxx) or a coiled-coil region, or an antibody directed against
CC these features. The vaccine is useful in protecting against infection
CC by Streptococcus pneumoniae. The vaccine composition comprising
CC antibodies to is useful for passive immunization for treating
CC pneumococcal infections which includes otitis media, nasopharyngeal
CC and bronchial infections.

XX Sequence 838 AA;

Query Match 7.5%; Score 60; DB 21; Length 838;
Best Local Similarity 100.0%; Pred. No. 1.2e-48;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHHYHYNGKVPYDAIISEELMKDPNY 90
Db 52 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHHYHYNGKVPYDAIISEELMKDPNY 111

RESULT 34

AAU75934
ID AAU75934 standard; Protein; 838 AA.

XX AAU75934;

XX 08-MAY-2002 (first entry)

XX Streptococcus pneumoniae BVH-11-2 protein.

XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;

XX pneumonia; streptococcal bacterial infection; BVH-11-2.

XX Streptococcus pneumoniae.

XX WO200198334-A2.

XX 27-DEC-2001.

XX 19-JUN-2001; 2001WO-CA00908.

XX 20-JUN-2000; 2000US-212683P.

XX (SHIR-) SHIRE BIOCHEM INC.

XX Hamel J, Queller C, Charland N, Martin D, Brodeur B;

XX MPI: 2002-122272/16.

XX N-PSDB; ABK15104.

XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and

XX epitope-bearing polypeptides, useful as vaccine components for treating

XX or preventing streptococcal infections such as otitis media,

XX meningitis, and bacteraemia

XX Example 1; Fig 8; 113pp; English.

XX The invention describes an isolated polypeptide (I) with 70-90%
XX identity to Streptococcus pneumoniae protein BVH-3, BVH-11, variants of
XX BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
XX comprising (I) is useful for therapeutic or prophylactic treatment of
XX meningitis, otitis media, bacteraemia or pneumonia infection in an
XX individual susceptible to these disorders. (II) is also useful for
XX therapeutic or prophylactic treatment of any streptococcal bacterial
XX infection (e.g., caused by Streptococcus pneumoniae, group A
XX Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
XX as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novae or
XX Streptococcus pneumoniae) in an individual susceptible to the infection.
XX A polynucleotide (III) encoding (I) is useful in DNA immunisation
XX techniques. The Streptococcus polypeptides are useful in a diagnostic
XX test for S. pneumoniae infection. (III) is useful for designing DNA
XX probes for use in detecting the presence of Streptococcus in a biological
XX sample suspected of containing the bacteria. The DNA probes may also be
XX used for detecting circulating S. pneumonia nucleic acid in a sample for
XX diagnosing streptococcal infections. This is the amino acid sequence of

CC Streptococcus pneumoniae protein BVH-11-2, used to create the antigenic
CC peptides described in the method of the invention.

XX
XX
SQ Sequence 838 AA;

Query Match 7.5%; Score 60; DB 23; Length 838;
Best Local Similarity 100.0%; Pred. No. 1.2e-48;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 31 ENLTPDEVSKREGINAEOIVIKITDQGYVTSBGDHYHYNGKVPYDAITSEBLMKDPNY 90
Db 52 ENLTPDEVSKREGINAEOIVIKITDQGYVTSBGDHYHYNGKVPYDAITSEBLMKDPNY 111

RESULT 35
ABU01418
ID ABU01418 standard; Protein; 839 AA.

XX
XX
AC ABU01418;
DT 11-FEB-2003 (first entry)

DE 5. pneumoniae type 4 strain protein from coding region #993.

XX Bacterial meningitis; pneumonia; sepsis; otitis media;
KW ear infection; antiinflammatory; antibacterial; immunostimulant;
KW auditory; respiratory; gene therapy; vaccine.

OS Streptococcus pneumoniae type 4 strain.

XX WO200277021-A2.

XX 03-OCT-2002.

XX 27-MAR-2002; 2002WO-IB02163.

XX 27-MAR-2001; 2001GB-0007658.

PA (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.

PI Masignani V, Tetteijn H, Fraser C;

DR MPI; 2003-040579/03.
DR N-PSDB; ABX06705.

PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT useful as medicaments for treating or preventing a disease or infection
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
PT or ear infection -

XX Claim 1, SEQ ID No 1986; 56pp; English.

XX The invention relates to a protein comprising or having at least 50%
XX identity to any of the 2469 amino acid sequences, identified in the
XX specification (available on a computer readable format), or its fragment,
XX expressed from 2469 of 2489 identified DNA coding regions from the
XX Streptococcus pneumoniae type 4 strain genomic sequence appearing as
XX AB556453. Also included are an antibody which binds one of the
XX proteins, treating a patient by administering the protein, DNA or
XX antibody (in a composition), a kit comprising first and second primers,
XX which are the nucleic acid cited above or fragments between nucleotides
XX 8-100 of a sequence not defined in the specification, for amplifying a
XX target sequence contained within a Streptococcus nucleic acid sequence,
XX where the first primer is substantially complementary to the target
XX sequence and the second primer is substantially complementary to the
XX complement of the target sequence, and where the parts of the primers
XX having substantial complementarity define the termini of the target
XX sequence to be amplified, assay comprising contacting a test compound
XX with the protein, and determining whether the test compound binds to the
XX protein and a Streptococcus pneumoniae bacterium, where one or more
XX genes encoding the proteins has been rendered inactive. The proteins,
XX nucleic acid molecules, antibody and compositions are useful as

CC medicaments for treating or preventing a disease or infection due to
CC streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,
CC sepsis, otitis media or ear infection. They are also useful in developing
CC vaccines, diagnostics and antibiotics. The methods are useful for
CC identifying immunodominant proteins. The present sequence is one of
CC the 2469 proteins expressed by the identified coding regions from the
CC genomic sequence.

CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 839 AA;

Query Match 7.5%; Score 60; DB 24; Length 839;
Best Local Similarity 100.0%; Pred. No. 1.2e-48;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 31 ENLTPDEVSKREGINAEOIVIKITDQGYVTSBGDHYHYNGKVPYDAITSEBLMKDPNY 90
Db 52 ENLTPDEVSKREGINAEOIVIKITDQGYVTSBGDHYHYNGKVPYDAITSEBLMKDPNY 111

RESULT 36
AAB12730
ID AAB12730 standard; Protein; 613 AA.

XX AAB12730;

DT 21-NOV-2000 (first entry)

DE Streptococcus pneumoniae BVH-11C protein antigen SEQ ID NO:63.

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-26; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.

OS Streptococcus pneumoniae.

XX WO2000039299-A2.

XX 06-JUL-2000.

XX 20-DEC-1999; 99WO-CA01218.

XX 23-DEC-1998; 98US-0113800.

PA (BIOC-) BIOCHEM PHARMA INC.

PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

DR MPI; 2000-452397/39.

PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia -

XX Claim 18; Fig 28; 106pp; English.

XX The present invention describes nucleic acids (I) encoding protein
XX antigens (II) from Streptococcus pneumoniae. The protein antigens
XX have bactericidal activity. The nucleic acids, encoding the protein
XX antigens, may be used for the recombinant production of the proteins
XX they encode. The protein antigens may then be used as vaccines for the
XX prevention and treatment of Streptococcal infections in mammals
XX (especially humans) which result in, e.g. meningitis, otitis media,
XX bacteraemia and/or pneumonia. The present sequence represents the
XX S. pneumoniae BVH-11C protein antigen.

XX Sequence 613 AA;

Query Match 7.0%; Score 56; DB 21; Length 613;
Best Local Similarity 100.0%; Pred. No. 6.7e-45;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 234 RTNWPSVSNPCTNTNTNTNSNTNSQASQSDNDIDSLKQLYKPLSQRHVESDGL 289
 DB 27 RTNWPSVSNPCTNTNTNTNSNTNSQASQSDNDIDSLKQLYKPLSQRHVESDGL 82

RESULT 37
 ID AAU84029 standard; Peptide; 613 AA.
 AC AAU84029;
 DT 08-MAY-2002 (first entry)
 DE Truncated variant of S. pneumoniae BVH-11, BVH- version #2.
 DE BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 KW pneumonia; streptococcal bacterial infection; mutant; mutein.
 OS Streptococcus pneumoniae.
 OS Synthetic.
 PN WO200198334-A2.
 PD 27-DEC-2001.
 PE 19-JUN-2001; 2001WO-CA00908.
 PR 20-JUN-2000; 2000US-212683P.
 PA (SHIR-) SHIRE BIOCHEM INC.
 PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 DR WPI: 2002-122272/16.
 PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia -
 PT
 PS Example 1; Page -, 113pp; English.
 CC The invention describes an isolated polypeptide (I) with 70-90%
 CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardiae or
 CC Staphylococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence represents a truncate
 CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
 CC described in the method of the invention.
 CC Note: This sequence does not appear in the specification but has
 CC been created according to information given in the invention.
 CC
 XX Sequence 613 AA;
 SQ

Query Match 7.0%; Score 56; DB 23; Length 613;
 Best Local Similarity 100.0%; Pred. No. 6,7e-45;
 Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 234 RTNWPSVSNPCTNTNTNTNSNTNSQASQSDNDIDSLKQLYKPLSQRHVESDGL 289

DB 27 RTNWPSVSNPCTNTNTNTNSNTNSQASQSDNDIDSLKQLYKPLSQRHVESDGL 82

RESULT 38
 ID AAY05753 standard; Protein; 721 AA.
 AC AAY05753;
 DT 19-JUL-1999 (first entry)
 DE Streptococcus pneumoniae 79 kDa complement C3-degrading protease.
 DE Human complement C3-degrading protease; vaccine; infection;
 KW meningitis; pneumonia; xerotransplantation; transplant rejection;
 KW inflammation.
 OS Streptococcus pneumoniae.
 OS
 FH Key Location/Qualifiers
 FT Protein 170..227
 FT /note="Claimed protein of Claim 14"
 FT Protein 258..300
 FT /note="Claimed protein of Claim 15"
 PN WO9915675-A1.
 PD 01-APR-1999.
 PE 24-SEP-1998; 98WO-US20186.
 PF 24-SEP-1997; 97US-0059907.
 PR 24-SEP-1997; 97US-0059907.
 PA (AMCY) AMERICAN CYANAMID CO.
 PA (MINU) UNIV MINNESOTA.
 PI Cheng Q, Finkel DJ, Green BA, Hostetter MK, Masi AW;
 DR WPI: 1999-254719/21.
 DR N-PSDB; AAX25394.
 XX
 PT New isolated human complement C3-degrading proteinase
 PS Claim 59; Page 54-58; 66pp; English.
 CC The present sequence is a 79 kDa protein of Streptococcus pneumoniae
 CC serotype 4 that is capable of degrading human complement protein C3
 CC (HCP3). The sequence was deduced from the open reading frame of
 CC an isolated gene (see AAX25394). A smaller (20 kDa) HCP3 protease
 CC is encoded by an open reading frame of S. pneumoniae serotype 3,
 CC and this open reading frame may be part of a larger open reading
 CC frame corresponding to the present sequence from the S. pneumoniae
 CC serotype 4 genome. Amino acids 1-58 and 90-132 of the 20 kDa
 CC protein have substantial sequence identity with amino acids 170-227
 CC and 258-300 of the 79 kDa protein. Proteins and peptides or
 CC polypeptides containing these regions, and DNA sequences encoding
 CC them, are claimed. HCP3 proteases and polypeptides can be used as
 CC immune system stimulating compositions (Claimed). They can produce
 CC an immune response against S. pneumoniae to immunize or treat a
 CC mammalian subject against infection or colonization (Claimed).
 CC They can produce a B cell response, a T cell response, an
 CC epithelial cell response, or an endothelial cell response
 CC (Claimed). The expression of the proteins on the surface of an
 CC organ of an animal used in xenotransplantation can be used to
 CC inhibit C3-mediated inflammation and rejection.
 CC
 XX Sequence 721 AA;
 SQ

Query Match 6.4%; Score 51; DB 20; Length 721;
 Best Local Similarity 100.0%; Pred. No. 5.2e-40;
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 QGRYTTDDGYIFNADIIEDTGDAYIVPHGDHYHYIPKNELASSELAAEA 210
 DB 74 QGRYTTDDGYIFNADIIEDTGDAYIVPHGDHYHYIPKNELASSELAAEA 124

RESULT 39

AAW55095
 ID AAW55095 standard; Protein; 763 AA.

AC AAW55095;

DT 02-OCT-1998 (first entry)

DE Streptococcus pneumoniae SP0042 protein.

KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
 detection; pneumonia; otitis media; meningitis.

OS Streptococcus pneumoniae.

Key Location/Qualifiers
 FT Misc-difference 51
 FT /label= unknown
 FT /note= "encoded by ATN"

FT Misc-difference 469
 FT /label= unknown
 FT /note= "encoded by NCA"

FT Misc-difference 477
 FT /label= unknown
 FT /note= "encoded by GAN"

PN WO9818930-A2.

PD 07-MAY-1998.

PF 30-OCT-1997; 97WO-US19422.

PR 31-OCT-1996; 96US-0023960.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Choi GH, Hromockyj A, Johnson LS, Kunsch CA;

DR WPI; 1998-272224/24.
 DR N-PSDB; AAV27356.

PT Nucleic acid encoding antigenic peptide(s) from Streptococcus
 pneumoniae - or their epitope-containing fragments, useful in
 protective or therapeutic vaccines, and for diagnosis

PS Claim 11; Page 62; 118pp; English.

CC The present sequence represents a protein from Streptococcus pneumoniae.
 CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein
 CC can be useful in vaccines for inducing protective antibodies against
 CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
 CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
 CC are used to detect Streptococcus infection (by usual hybridization or
 CC amplification methods), also for isolating Streptococcus genes or their
 CC allelic variants. The protein can be used similarly for detecting specific
 CC antibodies in standard immunoassays, especially for detecting specific
 CC monitoring infections. Antibodies which bind the protein are used to
 CC detect corresponding antigens, to purify the protein and for passive
 CC immunisation (optionally coupled to a toxin). Vaccines are administered,
 CC e.g. by injection, orally or through the skin, typically at 0.01-1000
 CC (especially 10-300) mu g/ml per dose.

SQ Sequence 763 AA;

Query Match 6.4%; Score 51; DB 19; Length 763;
 Best Local Similarity 100.0%; Pred. No. 5.5e-40;
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 QGRYTTDDGYIFNADIIEDTGDAYIVPHGDHYHYIPKNELASSELAAEA 210
 DB 159 QGRYTTDDGYIFNADIIEDTGDAYIVPHGDHYHYIPKNELASSELAAEA 209

RESULT 40

ABP54589

ID ABP54589 standard; Protein; 763 AA.

AC ABP54589;

DT 04-SEP-2002 (first entry)

DE S. pneumoniae SP042 protein sequence SEQ ID NO:66.

KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
 antibacterial; Streptococcal infection; detection.

OS Streptococcus pneumoniae.

PN US2002061545-A1.

PD 23-MAY-2002.

PF 22-JAN-2001; 2001US-0765272.

PR 30-OCT-1997; 97US-0961083.

PA (CHOI/) CHOI G H.

PA (KUNS/) KUNSCH C A.

PA (BARA/) BARASH S C.

PA (DILL/) DILLON P J.

PA (DOUG/) DOUGHERTY B.

PA (FANN/) FANNON M R.

PA (ROSE/) ROSEN C A.

PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;

DR WPI; 2002-479261/51.

DR N-PSDB; AB084824.

PT New Streptococcus pneumoniae antigens, useful for detecting

PT Streptococcus and for preventing or attenuating disease caused by

PT Streptococcus infection -

PS Claim 11; Page 29; 70pp; English.

CC AB084792 to AB084904 represents nucleic acids which encode the

CC Streptococcus pneumoniae antigens given in ABP54589 to ABP54669;

CC The S. pneumoniae antigens have antibacterial activity and can be

CC used in vaccines. The S. pneumoniae antigens can also be used to

CC prevent or attenuate a Streptococcal infection in an animal. The

CC polynucleotides encoding the S. pneumoniae antigens can be used to

CC detect Streptococcus nucleic acids. AB084905 to AB085130 represent

CC primers used in the cloning of S. pneumoniae ORFs (open reading frames)

CC which are used in an example from the present invention.

SQ Sequence 763 AA;

Query Match 6.4%; Score 51; DB 23; Length 763;

Best Local Similarity 100.0%; Pred. No. 5.5e-40;

Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 QGRYTTDDGYIFNADIIEDTGDAYIVPHGDHYHYIPKNELASSELAAEA 210

DB 159 QGRYTTDDGYIFNADIIEDTGDAYIVPHGDHYHYIPKNELASSELAAEA 209

RESULT 41

AAU83828

ID AAU83828 standard; Peptide; 42 AA.

```

AC  AAU83828;
XX
XX  08-MAY-2002 (first entry)
XX
XX  S. pneumoniae antigenic peptide 7G11.7.
XX
XX  BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
XX  pneumonia; streptococcal bacterial infection.
XX
XX  Streptococcus pneumoniae.
XX
XX  WO200198334-A2.
XX
XX  27-DEC-2001.
XX
XX  19-JUN-2001; 2001WO-CA00908.
XX
XX  20-JUN-2000; 2000US-212683P.
XX
XX  (SHIR-) SHIRE BIOCHEM INC.
XX
XX  Hamel J, Queller C, Charland N, Martin D, Brodeur B;
XX  WPI; 2002-122272/16.
XX
XX  New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX  epitope-bearing polypeptides, useful as vaccine components for treating
XX  or preventing streptococcal infections such as otitis media,
XX  meningitis, and bacteraemia.
XX
XX  Claim 10; Figure 12; 113pp; English.
XX
XX  The invention describes an isolated polypeptide (I) with 70-90%
XX  identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
XX  BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
XX  comprising (I) is useful for therapeutic or prophylactic treatment of
XX  meningitis, otitis media, bacteraemia or pneumonia infection in an
XX  individual susceptible to these disorders. (II) is also useful for
XX  therapeutic or prophylactic treatment of any streptococcal bacterial
XX  infection (e.g., caused by Streptococcus pneumoniae, group A
XX  Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
XX  as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novarcia or
XX  Streptococcus aureus) in an individual susceptible to the infection.
XX  A polynucleotide (III) encoding (I) is useful in DNA immunisation
XX  techniques. The Streptococcus polypeptides are useful in a diagnostic
XX  test for S. pneumoniae infection. (III) is useful for designing DNA
XX  probes for use in detecting the presence of Streptococcus in a biological
XX  sample suspected of containing the bacteria. The DNA probes may also be
XX  used for detecting circulating S. pneumonia nucleic acid in a sample for
XX  diagnosing streptococcal infections. This sequence represents an
XX  antigenic peptide derived from the S. pneumoniae genome, described in the
XX  method of the invention.
XX
XX  Sequence 42 AA;
XX
XX  Query Match 4.0%; Score 32; DB 23; Length 42;
XX  Best Local Similarity 100.0%; Pred. No. 8.7e-23;
XX  Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

XX
XX  Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
XX  prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX  otitis media; pneumonia; immunisation; bactericidal.
XX
XX  Streptococcus pneumoniae.
XX
XX  WO200039299-A2.
XX
XX  06-JUL-2000.
XX
XX  20-DEC-1999; 99WO-CA01218.
XX
XX  23-DEC-1998; 98US-0113800.
XX
XX  (BIOC-) BIOCHEM PHARMA INC.
XX
XX  Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX  WPI; 2000-452397/39.
XX
XX  Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX  otitis media, bacteraemia and/or pneumonia.
XX
XX  Claim 18; Fig 24; 106pp; English.
XX
XX  The present invention describes nucleic acids (I) encoding protein
XX  antigens (II) from Streptococcus pneumoniae. The protein antigens
XX  have bactericidal activity. The nucleic acids, encoding the protein
XX  antigens, may be used for the recombinant production of the proteins
XX  they encode. The protein antigens may then be used as vaccines for the
XX  prevention and treatment of Streptococcal infections in mammals
XX  (especially humans) which result in, e.g. meningitis, otitis media,
XX  bacteraemia and/or pneumonia. The present sequence represents the
XX  S. pneumoniae BVH-3C protein antigen.
XX
XX  Sequence 205 AA;
XX
XX  Query Match 4.0%; Score 32; DB 21; Length 205;
XX  Best Local Similarity 100.0%; Pred. No. 3.7e-22;
XX  Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY  46 AEOIVIKITDQGYVTSBGDHYHYNGKVPYDA 77
    |||||
DB  4 AEOIVIKITDQGYVTSBGDHYHYNGKVPYDA 35

```

```

XX
XX  RESULT 43
XX  ID AAU84025 standard; Peptide; 205 AA.
XX
XX  AAU84025;
XX
XX  08-MAY-2002 (first entry)
XX
XX  Truncated variant of S. pneumoniae BVH-3, BVH-3C.
XX
XX  BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
XX  pneumonia; streptococcal bacterial infection; mutant; mutein.
XX
XX  Streptococcus pneumoniae.
XX  Synthetic.
XX
XX  WO200198334-A2.
XX
XX  27-DEC-2001.
XX
XX  19-JUN-2001; 2001WO-CA00908.
XX
XX  20-JUN-2000; 2000US-212683P.
XX
XX  (SHIR-) SHIRE BIOCHEM INC.
XX

```

PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX
XX WPI; 2002-122272/16.
DR
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia
XX
XX
XX Example 1; Page -; 113pp; English.
XX
XX The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. necardia or
CC Streptococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a truncate
CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides.
CC described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
XX
SQ Sequence 205 AA;
Query Match 4.0%; Score 32; DB 23; Length 205;
Best Local Similarity 100.0%; Pred. No. 3.7e-22;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 46 AEOIVIKITDQGVTSHGPHYHYNGKVPYDA 77
DB 47 AEOIVIKITDQGVTSHGPHYHYNGKVPYDA 78
RESULT 44
ID AAW61228 standard; Protein; 447 AA.
XX
XX AAW61228;
AC
XX
XX 02-OCT-1998 (first entry)
DT
XX
XX Streptococcus pneumoniae SPI03 protein.
DE
XX
XX Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
KW detection; pneumonia; otitis media; meningitis.
XX
XX Streptococcus pneumoniae.
OS
XX
XX WO9818930-A2.
PN
XX
XX 07-MAY-1998.
PD
XX
XX 30-OCT-1997; 97WO-US19422.
PF
XX
XX 31-OCT-1996; 96US-0029960.
PR
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Choi GH, Hromockyj A, Johnson LS, Kunsch CA;
PI
XX
XX WPI; 1998-272224/24.

DR N-PSDB; AAV27414.
XX
XX Nucleic acid encoding antigenic peptide(s) from Streptococcus
PT pneumoniae - or their epitope-containing fragments, useful in
PT protective or therapeutic vaccines, and for diagnosis
XX
XX Claim 11; Page 85; 118pp; English.
XX
XX The present sequence represents a protein from Streptococcus pneumoniae.
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein
CC can be useful in vaccines for inducing protective antibodies against
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
CC are used to detect Streptococcus infection (by usual hybridisation or
CC amplification methods), also for isolating Streptococcus genes or their
CC allelic variants. The protein can be used similarly to detect specific
CC antibodies in standard immunoassays, especially for diagnosing or
CC monitoring infections. Antibodies which bind the protein are used to
CC detect corresponding antigens, to purify the protein and for passive
CC immunisation (optionally coupled to a toxin). Vaccines are administered,
CC e.g. by injection, orally or through the skin, typically at 0.01-1000
CC (especially 10-300) mg/g/ml per dose.
XX
SQ Sequence 447 AA;
Query Match 4.0%; Score 32; DB 19; Length 447;
Best Local Similarity 100.0%; Pred. No. 7.4e-22;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 46 AEOIVIKITDQGVTSHGPHYHYNGKVPYDA 77
DB 43 AEOIVIKITDQGVTSHGPHYHYNGKVPYDA 74
RESULT 45
ID ABP54647 standard; Protein; 447 AA.
XX
XX ABP54647;
AC
XX
XX 04-SEP-2002 (first entry)
DT
XX
XX S. pneumoniae SPI03 protein sequence SEQ ID NO:182.
DE
XX
XX Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
KW antibacterial; Streptococcal infection; detection.
XX
XX Streptococcus pneumoniae.
OS
XX
XX US2002061545-A1.
PN
XX
XX 23-MAY-2002.
PD
XX
XX 22-JAN-2001; 2001US-0765272.
PF
XX
XX 30-OCT-1997; 97US-0961083.
PR
XX
XX (CHOI/) CHOI G H.
PA (KUNS/) KUNSCH C A.
PA (BARA/) BARASH S C.
PA (DILL/) DILLON P J.
PA (DOUG/) DOUGHERTY B.
PA (FANN/) FANNON M R.
PA (ROSE/) ROSEN C A.
XX
XX Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
PI Rosen CA;
XX
XX WPI; 2002-479261/51.
DR N-PSDB; ABQ84882.
XX
XX New Streptococcus pneumoniae antigens, useful for detecting
PT Streptococcus and for preventing or attenuating disease caused by

PT Streptococcus infection -
 XX
 PS Claim 11; Page 45; 70pp; English.
 XX
 CC AB084972 to AB084904 represents nucleic acids which encode the
 CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669.
 CC The S. pneumoniae antigens have antibacterial activity and can be
 CC used in vaccines. The S. pneumoniae antigens can also be used to
 CC prevent or attenuate a Streptococcal infection in an animal. The
 CC polynucleotides encoding the S. pneumoniae antigens can be used to
 CC detect Streptococcus nucleic acids. AB084905 to AB085130 represent
 CC primers used in the cloning of S. pneumoniae ORFs (open reading frames)
 CC which are used in an example from the present invention.
 XX
 SQ Sequence 447 AA;
 Query Match 4.0%; Score 32; DB 23; Length 447;
 Best Local Similarity 100.0%; Pred. No. 7.4e-22;
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 46 AEOIVIKITDGGYVTSQDHYHYNGKVPYDA 77
 |||||
 Db 43 AEOIVIKITDGGYVTSQDHYHYNGKVPYDA 74
 RESULT 46
 AAB12718
 ID AAB12718 standard; Protein; 484 AA.
 AC AAB12718;
 XX
 DT 21-NOV-2000 (first entry)
 XX
 DE Streptococcus pneumoniae BVH-3A protein antigen SEQ ID NO:8.
 XX
 KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KW otitis media; pneumonia; immunisation; bactericidal.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200039299-A2.
 XX
 PD 06-JUL-2000.
 XX
 PF 20-DEC-1999; 99WO-CA01218.
 XX
 PR 23-DEC-1998; 98US-0113800.
 XX
 PA (BIOC-) BIOCHEM PHARMA INC.
 XX
 PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 XX
 DR WPI; 2000-452397/39.
 DR N-PSDB; AAA65733.
 XX
 PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteraemia and/or pneumonia -
 XX
 PS Claim 18; Fig 8; 106pp; English.
 XX
 CC The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the protein
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence represents the
 CC S. pneumoniae BVH-3A protein antigen.
 CC
 SQ Sequence 484 AA;

Query Match 4.0%; Score 32; DB 21; Length 484;
 Best Local Similarity 100.0%; Pred. No. 8e-22;
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 46 AEOIVIKITDGGYVTSQDHYHYNGKVPYDA 77
 |||||
 Db 67 AEOIVIKITDGGYVTSQDHYHYNGKVPYDA 98
 RESULT 47
 AAB01467
 ID AAB01467 standard; Protein; 484 AA.
 AC AAB01467;
 XX
 DT 20-OCT-2000 (first entry)
 XX
 DE Recombinant variant of Sp36 (Sp36E) of S. pneumoniae.
 XX
 KW Streptococcus pneumoniae; infection; vaccine; coiled coil region;
 KW histidine triad residue; Sp36; antibody; otitis media;
 KW nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
 KW meningitis; lobar pneumonia.
 XX
 OS Streptococcus pneumoniae.
 XX
 FH Key Location/Qualifiers
 FT Region 63..68
 FT /label= Histidine triad residue
 FT Region 120..140
 FT /label= Coiled coil region
 FT Region 185..190
 FT /label= Histidine triad residue
 FT Region 289..294
 FT /label= Histidine triad residue
 FT Region 376..381
 FT /label= Histidine triad residue
 FT Region 441..446
 FT /label= Histidine triad residue
 FT Region 750..772
 FT /label= Coiled coil region
 FT
 PN WO200037105-A2.
 XX
 PD 29-JUN-2000.
 XX
 PF 21-DEC-1999; 99WO-US30390.
 XX
 PR 21-DEC-1998; 98US-0113048.
 XX
 PA (MEDI-) MEDIMUNE INC.
 XX
 PI Johnson LS, Koenig S, Adamou JE;
 XX
 DR WPI; 2000-452129/39.
 DR N-PSDB; AAA47603.
 XX
 PT Vaccine useful for prophylaxis and treatment of pneumococcal infections
 PT such as otitis media, nasopharyngeal and bronchial infections,
 PT comprises Streptococcus pneumoniae proteins
 XX
 PS Claim 1; Page 58-60; 70pp; English.
 XX
 CC Although a number of proteins have been suggested as being involved
 CC in the pathogenicity of Streptococcus pneumoniae, there still remains
 CC a need to identify polypeptides having epitopes in common from
 CC various strains of S. pneumoniae in order to utilise such
 CC polypeptides in vaccines to protect against a wide variety of
 CC S. pneumoniae. New vaccine compositions are described which comprise a
 CC Streptococcus pneumoniae polypeptide (or fragments) of 80 - 680 amino
 CC acids in length that comprise at least one histidine triad residue
 CC (HxxHxH) or a coiled-coil region, or an antibody directed against
 CC these features. The vaccine is useful in protecting against infection

CC by *Streptococcus pneumoniae*. The vaccine composition comprising CC antitoxins is useful for passive immunization for treating CC pneumococcal infections which includes otitis media, nasopharyngeal CC and bronchial infections.

SQ Sequence 484 AA;

Query Match	4.0%	Score 32;	DB 21;	Length 484;
Best Local Similarity	100.0%;	Pred.No. 8e-22;		
Matches 32;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 46 AEQIVIKITDQGYVTSHGHHYHYNGKVPYDA 77
 Db 67 AEQIVIKITDQGYVTSHGHHYHYNGKVPYDA 98

RESULT 48
AAV81708
ID AAV81708 standard; Protein; 484 AA

AC AA81708

DT 02-JUN-2000 (first entry)

Streptococcus pneumoniae protein sequence ID128.

KM Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS
KM bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;
KM kidney disease; diabetes; immunosuppressive disorder; cistitis media;
KM pneumococcal septicemia; sinusitis; meningitis; therapy.

Streptococcus pneumoniae

PN WO200006738-A2.

PD 10-FEB-2000

PF 27-JUL-1999; 99WO-GB02452.

PR 27-JUL-1998; 98GB-0016336.

XX
2

2000

XX

DR N-PSDB; AAZ91804.

PT Streptococcal pro

XX Claim 1; Page 39; 76pp; English.
PS

CC This sequence represents a Streptococcus

iragments/are useful as immunogens or antigens. Immunogenic or antigenic compositions comprising the proteins are useful as vaccines and also in diagnostic assays. The sequences are useful for the detection or diagnosis of *S. pneumoniae* infection, by contacting a sample to be tested with them. Agents capable of antagonising, inhibiting or interfering with the function or expression of the protein or polypeptide are useful in medical compositions in the treatment of prophylaxis of *S. pneumoniae* infection. As the sequences can be used to treat *S. pneumoniae* infection, they can be used to treat bacterial pneumonia, which has high rates in young children, the elderly, and in patients with predisposing conditions such as asplenia, heart, lung and kidney disease, diabetes, alcoholism, or with immunosuppressive disorders, especially AIDS. They can also be used to treat pneumococcal septicaemia, otitis media, sinusitis, and meningitis.

SQ Sequence 484 AA;

Query Match	4.0%;	Score 32;	DB 21;	Length 484;
Best Local Similarity	100.0%;	Pred. No. 8e-22;		
Matches 32;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

46 AEQIVIKITDQGYVTSHGDDHYHYNGKVPYDA 77
67 AEQIVIKITDQGYVTSHGDDHYHYNGKVPYDA 98

ESULT 49
AY81538
D AY81538 standard; Protein; 485 AA

AAAY81538;

24-MAY-2000 (first entry)

Streptococcus pneumoniae type 4 protein sequence #38.

Streptococcus pneumoniae; vaccine; screening; protein antigen; antibacterial; anti-inflammatory; infection; diagnosis, pneumococcal disease.

Streptococcus pneumoniae.

WO200006737-A2

D 10-FEB-2000

F 27-JUL-1999; 99WO-GB02451

R 27-JUL-1998; 98GB-0016337

11. **X**

X

X

T New Streptococcal protein, useful as a vaccine, for diagnosis of
T Streptococcal diseases and for screening agents capable of antagonizing
T pneumococcal diseases -
T or inhibiting expression of the protein

Claim 1; Page 76; 108pp; English.

AAV815501 to AAV81679 represent specifically claimed protein sequences isolated from *Streptococcus pneumoniae*. AAV05407 to AAV05590 represent specifically claimed nucleotide sequences isolated from *S. pneumoniae*. The sequences have antibacterial and antiinflammatory properties. The protein sequences, and fragments of them, are useful as immunogens and/or antigens. The nucleotide sequences can be used in vaccines and in diagnostic assays. The proteins and nucleotides can be useful for the detection and diagnosis of *S. pneumoniae*. The protein sequences are also useful for screening an agent capable of antagonising, inhibiting or interfering with the function or expression of the proteins in which the agent is useful for treatment or prophylaxis of *S. pneumoniae* infection and meningitis. AAV05591 to AAV05614 represent primers used in the exemplification of the present invention.

Q Sequence 485 AA;

Query Match	4.0%	Score 32;	DB 21;	Length 485;
Best Local Similarity	100.0%	Pred. No. 8e-22;		
Matches 32;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

46 AEQIVIKITDQGYVTSHGHDHYHYNGKVPYDA 77
 67 AEQIVIKITDQGYVTSHGHDHYHYNGKVPYDA 98

RESULT 50
AB12723

```

ID AAB12723 standard; Protein; 489 AA.
XX
XX AAB12723;
AC
XX 21-NOV-2000 (first entry)
DT
XX Streptococcus pneumoniae BVH-3AD protein antigen SEQ ID NO:56.
DE
XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KM otitis media; pneumonia; immunisation; bactericidal.
XX
XX Streptococcus pneumoniae.
OS
XX WO200039299-A2.
PN
XX 06-JUL-2000.
PD
XX 20-DEC-1999; 99WO-CA01218.
PF
XX 23-DEC-1998; 98US-0113800.
PR
XX (BIOC-) BIOCHEM PHARMA INC.
PA
XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
PI WPI, 2000-452397/39.
XX
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
DR otitis media, bacteraemia and/or pneumonia -
XX
XX Claim 18; Fig 21; 106pp; English.
XX
XX The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents the
CC S. pneumoniae BVH-3AD protein antigen.
XX
XX
SQ Sequence 489 AA;
Query Match 4.0%; Score 32; DB 21; Length 489;
Best Local Similarity 100.0%; Pred. No. 8.1e-22;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 46 AEOIVIKITDGGYVTSHGDDHYHYNGKVPYDA 77
DB 47 AEOIVIKITDGGYVTSHGDDHYHYNGKVPYDA 78

```

```

XX
XX 19-JUN-2001; 2001WO-CA00908.
PF
XX 20-JUN-2000; 2000US-212683P.
XX
XX (SHIR-) SHIRE BIOCHEM INC.
PA
XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
PI WPI, 2002-122272/16.
XX
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia -
XX
XX Example 1; Page -; 113pp; English.
XX
XX The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. norcardia or
CC Streptococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a truncate
CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
CC described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
XX
XX
SQ Sequence 489 AA;
Query Match 4.0%; Score 32; DB 23; Length 489;
Best Local Similarity 100.0%; Pred. No. 8.1e-22;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 46 AEOIVIKITDGGYVTSHGDDHYHYNGKVPYDA 77
DB 47 AEOIVIKITDGGYVTSHGDDHYHYNGKVPYDA 78

```

```

RESULT 51
AAB12724
ID AAB12724 standard; Protein; 509 AA.
XX
XX AAB12724;
AC
XX 21-NOV-2000 (first entry)
DT
XX Streptococcus pneumoniae L-BVH-3-AD protein antigen SEQ ID NO:57.
DE
XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KM otitis media; pneumonia; immunisation; bactericidal.
XX
XX Streptococcus pneumoniae.
OS
XX WO200039299-A2.
PN
XX 06-JUL-2000.
PD
XX 20-DEC-1999; 99WO-CA01218.

```

```

RESULT 52
AAB12724
ID AAB12724 standard; Protein; 509 AA.
XX
XX AAB12724;
AC
XX 21-NOV-2000 (first entry)
DT
XX Streptococcus pneumoniae L-BVH-3-AD protein antigen SEQ ID NO:57.
DE
XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KM otitis media; pneumonia; immunisation; bactericidal.
XX
XX Streptococcus pneumoniae.
OS
XX WO200039299-A2.
PN
XX 06-JUL-2000.
PD
XX 20-DEC-1999; 99WO-CA01218.

```



```

PR 23-DEC-1996; 98US-0113800.
XX
XX (BIOC-) BIOCHEM PHARMA INC.
XX
XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
PI WPI; 2000-452397/39.
XX
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia -
XX
XX Claim 18; Fig 22; 106pp; English.
XX
XX The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents the
CC S. pneumoniae L-BVH-3-AD protein antigen.
XX
XX
SQ Sequence 509 AA;
XX
XX Query Match 4.0%; Score 32; DB 21; Length 509;
XX Best Local Similarity 100.0%; Pred. No. 8,4e-22;
XX Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 46 AEOIVIKITDQGYVTHSGDHYHYNGKVPYPA 77
XX ||||||||||||||||||||||||||||
XX 67 AEQIVKITDQGYVTHSGDHYHYNGKVPYDA 98
XX
XX
XX RESULT 53
XX AAU84023
XX ID AAU84023 standard; Peptide; 509 AA.
XX
XX AAU84023;
XX
XX 08-MAY-2002 (first entry)
XX
XX Truncated variant of S. pneumoniae BVH-3, L-BVH-3AD.
DE
XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutein.
XX
XX Streptococcus pneumoniae.
OS Synthetic.
XX
XX WO200198334-A2.
XX
XX 27-DEC-2001.
XX
XX 19-JUN-2001; 2001WO-CA00908.
XX
XX 20-JUN-2000; 2000US-212683P.
XX
XX (SHIR-) SHIRE BIOCHEM INC.
XX
XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
PI WPI; 2002-122272/16.
XX
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia -
XX
XX Example 1; Page -; 113pp; English.
XX
XX The invention describes an isolated polypeptide (I) with 70-90%
CC

```

CC	identify o Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC	BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC	comprising (I) is useful for therapeutic or prophylactic treatment of
CC	meningitis, otitis media, bacteraemia or pneumonia infection in an
CC	individual susceptible to these disorders. (II) is also useful for
CC	therapeutic or prophylactic treatment of any streptococcal bacterial
CC	infection (e.g., caused by Streptococcus pneumoniae, group A
CC	Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC	as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novacardia or
CC	Staphylococcus aureus) in an individual susceptible to the infection.
CC	A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC	techniques. The Streptococcus polypeptides are useful in a diagnostic
CC	test for S. pneumoniae infection. (III) is useful for designing DNA
CC	probes for use in detecting the presence of Streptococcus in a biological
CC	sample suspected of containing the bacteria. The DNA probes may also be
CC	used for detecting circulating S. pneumonia nucleic acid in a sample for
CC	diagnosing streptococcal infections. This sequence represents a truncate
CC	d of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
CC	described in the method of the invention.
CC	Note: This sequence does not appear in the specification but has
CC	been created according to information given in the invention.
XX	
XX	
SO	Sequence 509 AA;
OY	
Db	
	46 AEQIVIKITDGGYVTSHGDHYHYNGKVPDYA 77
	67 AEQIVIKITDGGYVTSHGDHYHYNGKVPDYA 98
RESULT 54	
AAU84092	
ID	AAU84092 standard; Peptide: 679 AA.
XX	
XX	AAU84092;
AC	
XX	
DT	08-MAY-2002 (first entry)
XX	
DE	Truncated variant of S. pneumoniae BVH-3, NEW15.
XX	
KW	BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
XX	pneumonia; streptococcal bacterial infection; mutant; mutein.
OS	Streptococcus pneumoniae.
XX	
FN	WO200198334-A2.
XX	
PD	27-DEC-2001.
PE	19-JUN-2001; 2001MO-CA00908.
XX	
PR	20-JUN-2000; 2000US-212683P.
XX	
PA	(SHIR-) SHIRE BIOCHEM INC.
PI	Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
DR	WPI; 2002-122272/16.
XX	
PT	New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT	epitope-bearing polypeptides, useful as vaccine components for treating
PT	or preventing streptococcal infections such as otitis media,
XX	meningitis, and bacteraemia -
PS	
XX	Example 1; Page -: 113bp; English.
CC	The invention describes an isolated polypeptide (I) with 70-90%
CC	identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC	BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC	comprising (I) is useful for therapeutic or prophylactic treatment of

CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (ii) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novae or
 CC Streptococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (iii) encoding (i) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (iii) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence represents a truncate
 CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
 CC described in the method of the invention.
 CC Note: This sequence does not appear in the specification but has
 CC been created according to information given in the invention.
 CC
 CC Sequence 679 AA;
 CC
 CC Query Match 4.0%; Score 32; DB 23; Length 679;
 CC Best Local Similarity 100.0%; Pred. No. 1.1e-21;
 CC Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 46 AEOIVIKITDGGYVTSQDHYHYNGKVPYDA 77
 CC |||||
 CC DB 47 AEOIVIKITDGGYVTSQDHYHYNGKVPYDA 78
 CC
 CC RESULT 55
 CC AAB12744
 CC ID AAB12744 standard; Protein: 780 AA.
 CC
 CC AC AAB12744;
 CC DT 21-NOV-2000 (first entry)
 CC XX
 CC DE Streptococcus pneumoniae NEM15 protein antigen SEQ ID NO:78.
 CC XX
 CC KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 CC KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 CC KW otitis media; pneumonia; immunisation; bactericidal.
 CC XX
 CC OS Streptococcus pneumoniae.
 CC XX
 CC FN WO200039299-A2.
 CC XX
 CC PD 06-JUL-2000.
 CC XX
 CC PF 20-DEC-1999; 99WO-CA01218.
 CC XX
 CC PR 23-DEC-1998; 98US-0113800.
 CC XX
 CC PA (BIOC-) BIOCHEM PHARMA INC.
 CC XX
 CC PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 CC XX
 CC DR WPI; 2000-452397/39.
 CC XX
 CC PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
 CC PT otitis media, bacteraemia and/or pneumonia -
 CC XX
 CC PS Claim 18; Fig 43; 106pp; English.
 CC
 CC The present invention describes nucleic acids (i) encoding protein
 CC antigens (ii) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the protein
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence represents the

CC S. pneumoniae NEM15 protein antigen.
 CC
 CC Sequence 780 AA;
 CC
 CC Query Match 4.0%; Score 32; DB 21; Length 780;
 CC Best Local Similarity 100.0%; Pred. No. 1.2e-21;
 CC Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 46 AEOIVIKITDGGYVTSQDHYHYNGKVPYDA 77
 CC |||||
 CC DB 47 AEOIVIKITDGGYVTSQDHYHYNGKVPYDA 78
 CC
 CC RESULT 56
 CC AAB12721
 CC ID AAB12721 standard; Protein: 840 AA.
 CC
 CC AC AAB12721;
 CC DT 21-NOV-2000 (first entry)
 CC XX
 CC DE Streptococcus pneumoniae SP63 BVH-3 protein antigen SEQ ID NO:16.
 CC XX
 CC KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 CC KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 CC KW otitis media; pneumonia; immunisation; bactericidal.
 CC XX
 CC OS Streptococcus pneumoniae.
 CC XX
 CC FN WO200039299-A2.
 CC XX
 CC PD 06-JUL-2000.
 CC XX
 CC PF 20-DEC-1999; 99WO-CA01218.
 CC XX
 CC PR 23-DEC-1998; 98US-0113800.
 CC XX
 CC PA (BIOC-) BIOCHEM PHARMA INC.
 CC XX
 CC PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 CC XX
 CC DR WPI; 2000-452397/39.
 CC XX
 CC DR N-PSDB; AAA65738.
 CC XX
 CC PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
 CC PT otitis media, bacteraemia and/or pneumonia -
 CC XX
 CC PS Claim 18; Fig 19; 106pp; English.
 CC
 CC The present invention describes nucleic acids (i) encoding protein
 CC antigens (ii) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the protein
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence represents the
 CC S. pneumoniae SP63 BVH-3 protein antigen.
 CC
 CC Sequence 840 AA;
 CC
 CC Query Match 4.0%; Score 32; DB 21; Length 840;
 CC Best Local Similarity 100.0%; Pred. No. 1.3e-21;
 CC Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 46 AEOIVIKITDGGYVTSQDHYHYNGKVPYDA 77
 CC |||||
 CC DB 47 AEOIVIKITDGGYVTSQDHYHYNGKVPYDA 78
 CC
 CC RESULT 57
 CC AAV76151
 CC ID AAV76151 standard; Protein: 840 AA.

XX	AAU76151;
AC	
XX	
DT	08-MAY-2002 (first entry)
XX	
DE	Streptococcus pneumoniae BVH-3 protein version #2.
XX	
KM	BVH-3; BVH-11, vaccine; meningitis; otitis media; bacteraemia;
XX	pneumonia; streptococcal bacterial infection.
OS	Streptococcus pneumoniae.
XX	
PN	WO200198334-A2.
XX	
PD	27-DEC-2001.
XX	
PF	19-JUN-2001; 2001WO-CA00908.
XX	
PR	20-JUN-2000; 2000US-212683P.
XX	
PA	(SHIR-) SHIRE BIOCHEM INC.
XX	
PI	Hamel J, Quellet C, Charla C, Charland N, Martin D, Brodeur B;
XX	
DR	WPI; 2002-122272/16.
XX	
XX	N-PSDB; ABR15105.
PT	New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT	epitope-bearing polypeptides, useful as vaccine components for treating
PT	or preventing streptococcal infections such as otitis media,
PT	meningitis, and bacteraemia
XX	
XX	
PS	Example 1; Fig 10; 113pp; English.
XX	
CC	The invention describes an isolated polypeptide (I) with 70-90%
CC	identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC	BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC	comprising (I) is useful for therapeutic or prophylactic treatment of
CC	meningitis, otitis media, bacteraemia or pneumonia infection in an
CC	individual susceptible to these disorders. (II) is also useful for
CC	therapeutic or prophylactic treatment of any streptococcal bacterial
CC	infection (e.g., caused by Streptococcus pneumoniae, group A
CC	Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC	as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novardi or
CC	Streptococcus aureus) in an individual susceptible to the infection.
CC	A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC	techniques. The Streptococcus polypeptides are useful in a diagnostic
CC	test for S. pneumoniae infection. (III) is useful for designing DNA
CC	probes for use in detecting the presence of Streptococcus in a biological
CC	sample suspected of containing the bacteria. The DNA probes may also be
CC	used for detecting circulating S. pneumonia nucleic acid in a sample for
CC	diagnosing streptococcal infections. This is the amino acid sequence of
CC	Streptococcus pneumoniae protein BVH-3, used to create the antigenic
CC	peptides described in the method of the invention.
CC	
XX	
XX	
SQ	Sequence 840 AA:
	Query Match 4.0%; Score 32; DB 23; Length 840;
	Best Local Similarity 100.0%; Pred. No. 1.3e-21;
	Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY	46 AEQIVIKITDQGYVTSHGDHYHYNGKVPYDA 77
Db	47 AEQIVIKITDQGYVTSHGDHYHYNGKVPYDA 78
RESULT 58	
ID	AAAB12722
XX	AAAB12722 standard; Protein; 1019 AA.
XX	
XX	AAAB12722;
XX	
DT	21-NOV-2000 (first entry)

DE	Streptococcus pneumoniae BVH-3M protein antigen SEQ ID NO:55.
XX	
XX	Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW	prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX	otitis media; pneumonia; immunisation; bactericidal.
OS	Streptococcus pneumoniae.
XX	
PN	WO200039299-A2.
PD	
XX	06-JUL-2000.
PE	
XX	20-DEC-1999; 99WO-COA01218.
PR	
XX	23-DEC-1998; 98US-0113800.
PA	(BIOC-) BIOCHEM PHARMA INC.
XX	
PI	Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N,
DR	WPI; 2000-452397/39.
PT	
XX	Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT	otitis media, bacteraemia and/or pneumonia -
PS	
XX	Claim 18; Fig 20; 106pp; English.
CC	
XX	The present invention describes nucleic acids (I) encoding protein
CC	antigens (II) from Streptococcus pneumoniae. The protein antigens
CC	have bactericidal activity. The nucleic acids, encoding the proteins
CC	antigens, may be used for the recombinant production of the proteins
CC	they encode. The protein antigens may then be used as vaccines for the
CC	prevention and treatment of Streptococcal infections in mammals
CC	(especially humans) which result in, e.g. meningitis, otitis media,
CC	bacteraemia and/or pneumonia. The present sequence represents the
CC	S. pneumoniae BVH-3M protein antigen.
XX	
SQ	Sequence 1019 AA:
	Query Match 4.0%; Score 32; DB 21; Length 1019;
	Best Local Similarity 100.0%; Pred. No. 1.6e-21;
	Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DQ	
	46 AEOIVIKITDGGVYTSHGDIHYHYNGKVPYPDA 77 47 AEOIVIKITDGGVYTSHGDIHYHYNGKVPYPDA 78
DB	
	RESULT 59
ID	AAB12748
AC	AAB12748 standard; Protein; 1019 AA.
XX	
AC	AAB12748;
XX	
DT	21-NOV-2000 (first entry)
DE	Streptococcus pneumoniae strain WU2 BVH-3 protein antigen.
XX	
KW	Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW	prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW	otitis media; pneumonia; immunisation; bactericidal.
OS	Streptococcus pneumoniae.
XX	
PN	WO200039299-A2.
PD	
XX	06-JUL-2000.
PF	
XX	20-DEC-1999; 99WO-COA01218.
XR	
XX	23-DEC-1998; 98US-0113800.
XX	

PA (BIOC-) BIOCHEM PHARMA INC.
 PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 XX WPI, 2000-452397/39.
 DR Streptococcal antigens useful for vaccinating against e.g. meningitis,
 XX otitis media, bacteremia and/or pneumonia -
 PT Disclosure; Fig 11; 106pp; English.
 XX
 CC The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the protein
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence represents a
 CC S. pneumoniae BVH-3 protein antigen, from the present invention.
 CC
 SQ Sequence 1019 AA;
 Query Match 4.0%; Score 32; DB 21; Length 1019;
 Best Local Similarity 100.0%; Pred. No. 1.6e-21;
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 46 AEOIVIKITDGGYVTSBGDHHYNGKVPYDA 77
 |||||||
 Db 47 AEOIVIKITDGGYVTSBGDHHYNGKVPYDA 78
 |||||||
 RESULT 60
 AAB12749
 ID AAB12749 standard; Protein, 1019 AA.
 XX
 AC AAB12749;
 XX
 DT 21-NOV-2000 (first entry)
 XX
 DE Streptococcus pneumoniae strain RX1 BVH-3 protein antigen.
 XX
 XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KM otitis media; pneumonia; immunisation; bactericidal.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200039299-A2.
 XX
 PD 06-JUL-2000.
 XX
 PF 20-DEC-1999; 99WO-CA01218.
 XX
 PR 23-DEC-1998; 98US-0113800.
 XX
 PA (BIOC-) BIOCHEM PHARMA INC.
 PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 XX WPI, 2000-452397/39.
 DR Streptococcal antigens useful for vaccinating against e.g. meningitis,
 XX otitis media, bacteremia and/or pneumonia -
 PT Disclosure; Fig 11; 106pp; English.
 XX
 CC The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the protein
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals

CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence represents a
 CC S. pneumoniae BVH-3 protein antigen, from the present invention.
 CC
 SQ Sequence 1019 AA;
 Query Match 4.0%; Score 32; DB 21; Length 1019;
 Best Local Similarity 100.0%; Pred. No. 1.6e-21;
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 46 AEOIVIKITDGGYVTSBGDHHYNGKVPYDA 77
 |||||||
 Db 47 AEOIVIKITDGGYVTSBGDHHYNGKVPYDA 78
 |||||||
 RESULT 61
 AAB12750
 ID AAB12750 standard; Protein, 1019 AA.
 XX
 AC AAB12750;
 XX
 DT 21-NOV-2000 (first entry)
 XX
 DE Streptococcus pneumoniae strain JNR7/87 BVH-3 protein antigen.
 XX
 XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KM otitis media; pneumonia; immunisation; bactericidal.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200039299-A2.
 XX
 PD 06-JUL-2000.
 XX
 PF 20-DEC-1999; 99WO-CA01218.
 XX
 PR 23-DEC-1998; 98US-0113800.
 XX
 PA (BIOC-) BIOCHEM PHARMA INC.
 PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 XX WPI, 2000-452397/39.
 DR Streptococcal antigens useful for vaccinating against e.g. meningitis,
 XX otitis media, bacteremia and/or pneumonia -
 PT Disclosure; Fig 11; 106pp; English.
 XX
 CC The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the protein
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence represents a
 CC S. pneumoniae BVH-3 protein antigen, from the present invention.
 CC
 SQ Sequence 1019 AA;
 Query Match 4.0%; Score 32; DB 21; Length 1019;
 Best Local Similarity 100.0%; Pred. No. 1.6e-21;
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 46 AEOIVIKITDGGYVTSBGDHHYNGKVPYDA 77
 |||||||
 Db 47 AEOIVIKITDGGYVTSBGDHHYNGKVPYDA 78
 |||||||
 RESULT 62
 AAB12751

```

ID  AAB12751 standard; Protein; 1019 AA.
XX
XX  AAB12751;
AC
XX
XX  21-NOV-2000 (first entry)
DT
XX
XX  Streptococcus pneumoniae strain SP64 BVH-3 protein antigen.
DE
XX
XX  Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW  prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX  otitis media; pneumonia; immunisation; bactericidal.
XX
OS  Streptococcus pneumoniae.
XX
XX  WO200039299-A2.
XX
XX  06-JUL-2000.
XX
XX  20-DEC-1999; 99WO-CA01218.
XX
XX  23-DEC-1998; 98US-0113800.
XX
XX  (BIOC-) BIOCHEM PHARMA INC.
XX
XX  Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
PI
XX  WPI; 2000-452397/39.
XX
XX  Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT  otitis media, bacteraemia and/or pneumonia -
XX
XX  Disclosure; Fig 11; 106pp; English.
XX
XX
XX  The present invention describes nucleic acids (I) encoding protein
CC  antigens (II) from Streptococcus pneumoniae. The protein antigens
CC  have bactericidal activity. The nucleic acids, encoding the protein
CC  antigens, may be used for the recombinant production of the proteins
CC  they encode. The protein antigens may then be used as vaccines for the
CC  prevention and treatment of Streptococcal infections in mammals
CC  (especially humans) which result in, e.g. meningitis, otitis media,
CC  bacteraemia and/or pneumonia. The present sequence represents a
CC  S. pneumoniae BVH-3 protein antigen, from the present invention.
XX
XX  Sequence 1019 AA;
SQ
XX
XX  Query Match 4.0%; Score 32; DB 21; Length 1019;
XX  Best Local Similarity 100.0%; Pred. No. 1.6e-21;
XX  Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX  46 AEOIVIKITDQGYVTSBGDHYHYNGKVPYDA 77
OY  |||||||||||||||||||||||||||||||||||
DB  47 AEOIVIKITDQGYVTSBGDHYHYNGKVPYDA 78

RESULT 63
AAB12752
ID  AAB12752 standard; Protein; 1019 AA.
XX
XX  AAB12752;
AC
XX
XX  21-NOV-2000 (first entry)
DT
XX
XX  Streptococcus pneumoniae strain P4241 BVH-3 protein antigen.
DE
XX
XX  Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW  prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX  otitis media; pneumonia; immunisation; bactericidal.
XX
OS  Streptococcus pneumoniae.
XX
XX  WO200039299-A2.
XX
XX  06-JUL-2000.
XX

```

```

XX
XX  20-DEC-1999; 99WO-CA01218.
PF
XX
XX  23-DEC-1998; 98US-0113800.
PR
XX
XX  (BIOC-) BIOCHEM PHARMA INC.
XX
XX  Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
PI
XX  WPI; 2000-452397/39.
XX
XX  Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT  otitis media, bacteraemia and/or pneumonia -
XX
XX  Disclosure; Fig 11; 106pp; English.
XX
XX
XX  The present invention describes nucleic acids (I) encoding protein
CC  antigens (II) from Streptococcus pneumoniae. The protein antigens
CC  have bactericidal activity. The nucleic acids, encoding the protein
CC  antigens, may be used for the recombinant production of the proteins
CC  they encode. The protein antigens may then be used as vaccines for the
CC  prevention and treatment of Streptococcal infections in mammals
CC  (especially humans) which result in, e.g. meningitis, otitis media,
CC  bacteraemia and/or pneumonia. The present sequence represents a
CC  S. pneumoniae BVH-3 protein antigen, from the present invention.
XX
XX  Sequence 1019 AA;
SQ
XX
XX  Query Match 4.0%; Score 32; DB 21; Length 1019;
XX  Best Local Similarity 100.0%; Pred. No. 1.6e-21;
XX  Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX  46 AEOIVIKITDQGYVTSBGDHYHYNGKVPYDA 77
OY  |||||||||||||||||||||||||||||||||||
DB  47 AEOIVIKITDQGYVTSBGDHYHYNGKVPYDA 78

RESULT 64
AAB12753
ID  AAB12753 standard; Protein; 1019 AA.
XX
XX  AAB12753;
AC
XX
XX  21-NOV-2000 (first entry)
DT
XX
XX  Streptococcus pneumoniae strain A66 BVH-3 protein antigen.
DE
XX
XX  Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW  prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX  otitis media; pneumonia; immunisation; bactericidal.
XX
OS  Streptococcus pneumoniae.
XX
XX  WO200039299-A2.
XX
XX  06-JUL-2000.
XX
XX  20-DEC-1999; 99WO-CA01218.
XX
XX  23-DEC-1998; 98US-0113800.
XX
XX  (BIOC-) BIOCHEM PHARMA INC.
XX
XX  Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
PI
XX  WPI; 2000-452397/39.
XX
XX  Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT  otitis media, bacteraemia and/or pneumonia -
XX
XX  Disclosure; Fig 11; 106pp; English.
XX
XX  The present invention describes nucleic acids (I) encoding protein
CC

```

CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the protein
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence represents a
 CC S. pneumoniae BVH-3 protein antigen, from the present invention.

XX Sequence 1019 AA;

Query Match 4.0%; Score 32; DB 21; Length 1019;

Best Local Similarity 100.0%; Pred. No. 1.6e-21;
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 46 AEQIVIKITDGYVTSKGDHYYNGKVPYDA 77
 |||||
 Db 47 AEQIVIKITDGYVTSKGDHYYNGKVPYDA 78

RESULT 65

AAU84021
 ID AAU84021 standard; Peptide; 1019 AA.

XX AAU84021;

DT 08-MAY-2002 (first entry)

DE Truncated variant of S. pneumoniae BVH-3, BVH-3M.

KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 KM pneumonia; streptococcal bacterial infection; mutant; mucin.

XX Streptococcus pneumoniae.

OS Synthetic.

PN WO200198334-A2.

XX 27-DEC-2001.

PF 19-JUN-2001; 2001WO-CA00908.

PR 20-JUN-2000; 2000US-212683P.

XX (SHR-) SHIRE BIOCHEM INC.

PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

DR WPI; 2002-122272/16.

PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia

PS Example 1; Page -, 113pp; English.

XX The invention describes an isolated polypeptide (I) with 70-90%
 CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novardi or
 CC Streptococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful for designing DNA
 CC test for S. pneumoniae infection. (II) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be

CC used for detecting circulating S. pneumonia nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence represents a truncate
 CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
 CC described in the method of the invention.
 CC Note: This sequence does not appear in the specification but has
 CC been created according to information given in the invention.

XX Sequence 1019 AA;

Query Match 4.0%; Score 32; DB 23; Length 1019;

Best Local Similarity 100.0%; Pred. No. 1.6e-21;
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 46 AEQIVIKITDGYVTSKGDHYYNGKVPYDA 77
 |||||
 Db 47 AEQIVIKITDGYVTSKGDHYYNGKVPYDA 78

RESULT 66

AAB12715
 ID AAB12715 standard; Protein; 1039 AA.

XX AAB12715;

DT 21-NOV-2000 (first entry)

DE Streptococcus pneumoniae BVH-3 protein antigen SEQ ID NO:2.

KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KM prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KM otitis media; pneumonia; immunisation; bactericidal.

XX Streptococcus pneumoniae.

OS WO200039299-A2.

PN 06-JUL-2000.

XX 20-DEC-1999; 99WO-CA01218.

PR 23-DEC-1998; 98US-0113800.

XX (BIOC-) BIOCHEM PHARMA INC.

PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

DR WPI; 2000-452397/39.

DR N-PSDB; AAA65730.

PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteraemia and/or pneumonia

PS Claim 18; Fig 2; 106pp; English.

XX The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the proteins
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence represents the
 CC S. pneumoniae BVH-3 protein antigen.

XX Sequence 1039 AA;

Query Match 4.0%; Score 32; DB 21; Length 1039;

Best Local Similarity 100.0%; Pred. No. 1.6e-21;
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 46 AEQIVIKITDGYVTSKGDHYYNGKVPYDA 77
 |||||
 Db 67 AEQIVIKITDGYVTSKGDHYYNGKVPYDA 98

RESULT 67

AAU75932
ID AAU75932 standard; Protein; 1039 AA.

AAU75932;

08-MAY-2002 (first entry)

Streptococcus pneumoniae BVH-3 protein version #1.

BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
pneumonia; streptococcal bacterial infection.

Streptococcus pneumoniae.

WO200198334-A2.

27-DEC-2001.

19-JUN-2001; 2001WO-CA00908.

20-JUN-2000; 2000US-212683P.

(SHIR-) SHIRE BIOCHEM INC.

Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

WPI; 2002-122272/16.

N-PSDB; ABK15101.

New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
epitope-bearing polypeptides, useful as vaccine components for treating
or preventing streptococcal infections such as otitis media,
meningitis, and bacteraemia

Example 1; Fig 6; 113pp; English.

The invention describes an isolated polypeptide (I) with 70-90%
identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
comprising (I) is useful for therapeutic or prophylactic treatment of
meningitis, otitis media, bacteraemia or pneumonia infection in an
individual susceptible to these disorders. (II) is also useful for
therapeutic or prophylactic treatment of any streptococcal bacterial
infection (e.g., caused by Streptococcus pneumoniae, group A
Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novacardia or
Staphylococcus aureus) in an individual susceptible to the infection.
A polynucleotide (III) encoding (I) is useful in DNA immunisation
techniques. The Streptococcus polypeptides are useful in a diagnostic
test for S. pneumoniae infection. (III) is useful for designing DNA
probes for use in detecting the presence of Streptococcus in a biological
sample suspected of containing the bacteria. The DNA probes may also be
used for detecting circulating S. pneumonia nucleic acid in a sample for
diagnosing streptococcal infections. This is the amino acid sequence of
Streptococcus pneumoniae protein BVH-3, used to create the antigenic
peptides described in the method of the invention.

Sequence 1039 AA;

Query Match 4.0%; Score 32; DB 23; Length 1039;

Best Local Similarity 100.0%; Pred. No. 1.6e-21;

Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

46 AEOIVIKITDQGVYTSHGDIHYHYNGKVPYDA 77

67 AEOIVIKITDQGVYTSHGDIHYHYNGKVPYDA 98

RESULT 68

ABU01419

ABU01419 standard; Protein; 1039 AA.

ABU01419;

11-FEB-2003 (first entry)

S. pneumoniae type 4 strain protein from coding region #994.

Bacterial meningitis; pneumonia; sepsis; otitis media;

ear infection; antiinflammatory; antibacterial; immunostimulant;

auditory; respiratory; gene therapy; vaccine.

Streptococcus pneumoniae type 4 strain.

WO200277021-A2.

03-OCT-2002.

27-MAR-2002; 2002WO-1B02163.

27-MAR-2001; 2001GB-0007658.

(CHIR-) CHIRON SPA.

(GENO-) INST GENOMIC RES.

Masignani V, Tettelin H, Fraser C;

WPI; 2003-040579/03.

N-PSDB; ABX06706.

New proteins and nucleic acid molecules from Streptococcus pneumoniae,
useful as medicaments for treating or preventing a disease or infection
due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
or ear infection

Claim 1; SEQ ID No 1988; 56pp; English.

The invention relates to a protein comprising or having at least 50%
identity to any of the 2469 amino acid sequences, identified in the
specification (available on a computer readable format), or its fragment,
expressed from 2469 of 2489 identified DNA coding regions from the
Streptococcus pneumoniae type 4 strain genomic sequence appearing as
AB56454. Also included are an antibody which binds one of the
proteins, treating a patient by administering the protein, DNA or
antibody (in a composition), a kit comprising first and second primers,
which are the nucleic acid cited above or fragments between nucleotides
8-100 of a sequence not defined in the specification, for amplifying a
target sequence contained within a Streptococcus nucleic acid sequence,
where the first primer is substantially complementary to the target
sequence and the second primer is substantially complementary to the
complement of the target sequence, and where the parts of the primers
having substantial complementarity define the termini of the target
sequence to be amplified, assay comprising contacting a test compound
with the protein, and determining whether the test compound binds to the
protein and a Streptococcus pneumoniae bacterium, where one or more
genes encoding the proteins has been rendered inactive. The proteins,
nucleic acid molecules, antibody and compositions are useful as
medicaments for treating or preventing a disease or infection due to
Streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,
sepsis, otitis media or ear infection. They are also useful in developing
vaccines, diagnostics and antibiotics. The methods are useful for
identifying immunodominant proteins. The present sequence is one of
the 2469 proteins expressed by the identified coding regions from the
genomic sequence.
Note: The sequence data for this patent did not form part
of the printed specification, but was obtained in electronic
format directly from WIPO at
ftp.wipo.int/pub/published_pct_sequences.

Sequence 1039 AA;

Query Match 4.0%; Score 32; DB 24; Length 1039;

Best Local Similarity 100.0%; Pred. No. 1.6e-21;

Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 46 AEOIVIKITDQGYTSHGDHYHYNGKVPYDA 77
 DB 67 AEOIVIKITDQGYTSHGDHYHYNGKVPYDA 98

RESULT 69

ID AUB4094 standard; Peptide; 294 AA.

AC AUB4094;

DT 08-MAY-2002 (first entry)

DE Truncated variant of S. pneumoniae BVH-11-2, NEM18.

KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;

KW pneumoniae; streptococcal bacterial infection; mutant; mulein.

OS Streptococcus pneumoniae.

PN WO200198334-A2.

PD 27-DEC-2001.

PF 19-JUN-2001; 2001WO-CA00908.

PR 20-JUN-2000; 2000US-212683P.

XX (SHIR-) SHIRE BIOCHEM INC.

PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

DR WPI; 2002-122272/16.

XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and

PT epitope-bearing polypeptides, useful as vaccine components for treating

PT or preventing streptococcal infections such as otitis media,

PT meningitis, and bacteraemia -

PS Example 1; Page -; 113pp; English.

XX The invention describes an isolated polypeptide (I) with 70-90%

CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of

CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)

CC comprising (I) is useful for therapeutic or prophylactic treatment of

CC meningitis, otitis media, bacteraemia or pneumonia infection in an

CC individual susceptible to these disorders. (II) is also useful for

CC therapeutic or prophylactic treatment of any streptococcal bacterial

CC infection (e.g., caused by Streptococcus pneumoniae, group A

CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such

CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nodocidia or

CC Streptococcus aureus) in an individual susceptible to the infection.

CC A polynucleotide (III) encoding (I) is useful in DNA immunisation

CC techniques. The Streptococcus polypeptides are useful in a diagnostic

CC test for S. pneumoniae infection. (III) is useful for designing DNA

CC probes for use in detecting the presence of Streptococcus in a biological

CC sample suspected of containing the bacteria. The DNA probes may also be

CC used for detecting circulating S. pneumonia nucleic acid in a sample for

CC diagnosing streptococcal infections. This sequence represents a truncate

CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,

CC described in the method of the invention.

CC Note: This sequence does not appear in the specification but has

CC been created according to information given in the invention.

CC

XX

SQ Sequence 294 AA;

Query Match 3.4%; Score 27; DB 23; Length 294;

Best Local Similarity 100.0%; Pred. No. 3.4e-17;

Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 326 RIARILPLRYSNMHWVPSRPEQSPQ 352

DB 106 RIARILPLRYSNMHWVPSRPEQSPQ 132

RESULT 70

ID ABP26208 standard; Protein; 466 AA.

AC ABP26208;

DT 02-JUL-2002 (first entry)

DE Streptococcus polypeptide SEQ ID NO 1592.

KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;

KW group A streptococcus; Streptococcus pyogenes; antibacterial;

XX antiinflammatory; infection; vaccine; meningitis; gene therapy.

OS Streptococcus agalactiae.

PN WO200234771-A2.

PD 02-MAY-2002.

PF 29-OCT-2001; 2001WO-GB04789.

PR 27-OCT-2000; 2000GB-0026333.

PR 24-NOV-2000; 2000GB-0028727.

PR 07-MAR-2001; 2001GB-0005640.

XX (CHIR-) CHIRON SPA.

PA (GENO-) INST GENOMIC RES.

PI Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;

PI Tectelin H;

DR WPI; 2002-352536/38.

DR N-PSDB; ABN66839.

XX New Streptococcus protein for the treatment or prevention of infection

PT or disease caused by Streptococcus bacteria, such as meningitis, and

PT for detecting a compound that binds to the protein -

PS Claim 1; Page 3310; 4525pp; English.

XX The invention relates to a protein (ABP25413-ABP30895) from group B

CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS

CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in

CC the specification. The proteins have antibacterial and antiinflammatory

CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and

CC antibodies that bind (I) are used in the manufacture of medicaments for

CC the treatment or prevention of infection or disease caused by

CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.

CC Nucleic acids encoding (I) are used to detect Streptococcus in a

CC biological sample. (I) is used to determine whether a compound binds to

CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be

CC used as a vaccine or diagnostic composition. The disease caused by

CC Streptococcus that is prevented or treated may be meningitis. Nucleic

CC acid encoding (I) may be used to recombinantly produce (I) and may be

CC used in gene therapy. Antibodies to (I) are used for affinity

CC chromatography, immunoassays, and distinguishing/identifying

CC Streptococcus proteins.

XX

SQ Sequence 466 AA;

Query Match 3.4%; Score 27; DB 23; Length 466;

Best Local Similarity 100.0%; Pred. No. 5.2e-17;

Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 57 GYVTSBGDHYHYNGKVPYDAIISEEL 83

DB 81 GYVTSBGDHYHYNGKVPYDAIISEEL 107


```
RESULT 71
AAB12743
ID AAB12743 standard; Protein; 473 AA.
XX
AC AAB12743;
XX
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae NEW12 protein antigen SEQ ID NO:77.
XX
KM Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX ctitis media; pneumonia; immunisation; bactericidal.
XX
OS Streptococcus pneumoniae.
XX
PN WO200039299-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA01218.
XX
PR 23-DEC-1998; 98US-0113800.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
XX
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
DR WPI; 2000-452397/39.
XX
PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX ctitis media, bacteraemia and/or pneumonia -
XX
PS Claim 18; Fig 42; 106pp; English.
XX
CC The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, ctitis media,
CC bacteraemia and/or pneumonia. The present sequence represents the
CC S. pneumoniae NEW14 protein antigen.
XX
SQ Sequence 473 AA;
XX
Query Match 3.4%; Score 27; DB 21; Length 473;
Best Local Similarity 100.0%; Pred. No. 5.3e-17;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 326 RIRRIIPLRYSNHWVDSRPEQSPQ 352
DB 106 RIRRIIPLRYSNHWVDSRPEQSPQ 132
XX
RESULT 72
AAB12741
ID AAB12741 standard; Peptide; 473 AA.
XX
AC AAB12741;
XX
DT 08-MAY-2002 (first entry)
XX
DE Truncated variant of S. pneumoniae BVH-11-2, NEW14.
XX
KM BVH-3; BVH-11; vaccine; meningitis; ctitis media; bacteraemia;
XX pneumonia; streptococcal bacterial infection; mutant; mtein.
XX
OS Streptococcus pneumoniae.
XX
PN WO200198334-A2.
XX
```

```
XX
PD 27-DEC-2001.
XX
PF 19-JUN-2001; 2001WO-CA00908.
XX
PR 20-JUN-2000; 2000US-212683P.
XX
PA (SHIR-) SHIRE BIOCHEM INC.
XX
PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
DR WPI; 2002-122272/16.
XX
PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as ctitis media,
XX meningitis, and bacteraemia -
XX
PS Example 1; Page -, 113pp; English.
XX
CC The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, ctitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novacardia or
CC Staphylococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a truncate
CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
CC described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
XX
SQ Sequence 473 AA;
XX
Query Match 3.4%; Score 27; DB 21; Length 473;
Best Local Similarity 100.0%; Pred. No. 5.3e-17;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 326 RIRRIIPLRYSNHWVDSRPEQSPQ 352
DB 106 RIRRIIPLRYSNHWVDSRPEQSPQ 132
XX
RESULT 73
AAB12741
ID AAB12741 standard; Protein; 568 AA.
XX
AC AAB12741;
XX
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae NEW10 protein antigen SEQ ID NO:74.
XX
KM Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX ctitis media; pneumonia; immunisation; bactericidal.
XX
OS Streptococcus pneumoniae.
XX
PN WO200039299-A2.
XX
PD 06-JUL-2000.
XX
```

XX 20-DEC-1999; 99WO-CA01218.
 XX 23-DEC-1998; 98US-0113800.
 XX (BIOC-) BIOCHEM PHARMA INC.
 XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 XX WPI; 2000-452397/39.
 XX
 XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteraemia and/or pneumonia -
 XX
 XX Claim 18; Fig 39; 106pp; English.
 XX
 XX The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the protein
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence represents the
 CC S. pneumoniae NEW10 protein antigen.
 XX
 XX Sequence 568 AA;
 SQ

Query Match 3.4%; Score 27; DB 21; Length 568;
 Best Local Similarity 100.0%; Pred. No. 6,2e-17;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 326 RIARIIPLRYSNHWVPDSRPEQSPQ 352
 |||||||
 DB 62 RIARIIPLRYSNHWVPDSRPEQSPQ 88

RESULT 74
 AAU84088
 ID AAU84088 standard; Peptide; 568 AA.
 XX
 AC AAU84088;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE Truncated variant of S. pneumoniae BVH-11-2, NEW10.
 XX
 KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 KW pneumonia; streptococcal bacterial infection; mutant; mutein.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200198334-A2.
 XX
 PD 27-DEC-2001.
 XX
 PF 19-JUN-2001; 2001WO-CA00908.
 XX
 PR 20-JUN-2000; 2000US-212683P.
 XX
 PA (SHIR-) SHIRE BIOCHEM INC.
 XX
 PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 XX
 DR WPI; 2002-122272/16.
 XX
 PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia -
 XX
 XX Example 1; Page -; 113pp; English.
 XX

CC The invention describes an isolated polypeptide (I) with 70-90%
 CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. norcardia or
 CC Streptococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence represents a truncate
 CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
 CC described in the method of the invention.
 CC Note: This sequence does not appear in the specification but has
 CC been created according to information given in the invention.
 XX
 XX Sequence 568 AA;
 SQ

Query Match 3.4%; Score 27; DB 23; Length 568;
 Best Local Similarity 100.0%; Pred. No. 6,2e-17;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 326 RIARIIPLRYSNHWVPDSRPEQSPQ 352
 |||||||
 DB 62 RIARIIPLRYSNHWVPDSRPEQSPQ 88

RESULT 75
 AAU84031
 ID AAU84031 standard; Peptide; 612 AA.
 XX
 AC AAU84031;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE Truncated variant of S. pneumoniae BVH-11, NEW24.
 XX
 KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 KW pneumonia; streptococcal bacterial infection; mutant; mutein.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200198334-A2.
 XX
 PD 27-DEC-2001.
 XX
 PF 19-JUN-2001; 2001WO-CA00908.
 XX
 PR 20-JUN-2000; 2000US-212683P.
 XX
 PA (SHIR-) SHIRE BIOCHEM INC.
 XX
 PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 XX
 DR WPI; 2002-122272/16.
 XX
 PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia -
 XX
 XX Example 1; Page -; 113pp; English.
 XX
 CC The invention describes an isolated polypeptide (I) with 70-90%
 CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of

BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II) comprising (I) is useful for therapeutic or prophylactic treatment of meningitis, otitis media, bacteraemia or pneumonia infection in an individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by Streptococcus pneumoniae, group A Streptococcus such as Streptococcus pyogenes, group B Streptococcus such as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardiae or Streptococcus aureus) in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The Streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (III) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating S. pneumonia nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a truncate of a Streptococcus pneumoniae gene used to obtain antigenic peptides, described in the method of the invention.

Note: This sequence does not appear in the specification but has been created according to information given in the invention.

Sequence 612 AA;

Query Match 3.4%; Score 27; DB 23; Length 612;
Best Local Similarity 100.0%; Pred. No. 6.6e-17;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

326 RIARIIPLRYSNMHWPDSPRPEQSPQ 352
106 RIARIIPLRYSNMHWPDSPRPEQSPQ 132

RESULT 76

AAU84055 standard; Peptide; 1139 AA.

AAU84055;

08-MAY-2002 (first entry)

S. pneumoniae derived chimeric peptide, NEW28.

BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
pneumonia; streptococcal bacterial infection; mutant; murein;
BVH-11-2.

Streptococcus pneumoniae.
Synthetic.

WO200198334-A2.

27-DEC-2001.

19-JUN-2001; 2001WO-CA00908.

20-JUN-2000; 2000US-212683P.

(SHIR-) SHIRE BIOCHEM INC.

Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

WPI; 2002-122272/16.

New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
epitope-bearing polypeptides, useful as vaccine components for treating
or preventing streptococcal infections such as otitis media,
meningitis, and bacteraemia

Example 1; Page -: 113pp; English.

The invention describes an isolated polypeptide (I) with 70-90%
identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)

comprising (I) is useful for therapeutic or prophylactic treatment of meningitis, otitis media, bacteraemia or pneumonia infection in an individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by Streptococcus pneumoniae, group A Streptococcus such as Streptococcus pyogenes, group B Streptococcus such as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardiae or Streptococcus aureus) in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The Streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (III) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating S. pneumonia nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a chimeric gene created from fragments and variant fragments of Streptococcus pneumoniae genes, described in the method of the invention.

Note: This sequence does not appear in the specification but has been created according to information given in the invention.

Sequence 1139 AA;

Query Match 3.4%; Score 27; DB 23; Length 1139;
Best Local Similarity 100.0%; Pred. No. 1.2e-16;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

326 RIARIIPLRYSNMHWPDSPRPEQSPQ 352
63 RIARIIPLRYSNMHWPDSPRPEQSPQ 89

RESULT 77

AAU84053 standard; Peptide; 1378 AA.

AAU84053;

08-MAY-2002 (first entry)

S. pneumoniae derived chimeric peptide, NEW26.

BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
pneumonia; streptococcal bacterial infection; mutant; murein;
BVH-11-2.

Streptococcus pneumoniae.
Synthetic.

WO200198334-A2.

27-DEC-2001.

19-JUN-2001; 2001WO-CA00908.

20-JUN-2000; 2000US-212683P.

(SHIR-) SHIRE BIOCHEM INC.

Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

WPI; 2002-122272/16.

New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
epitope-bearing polypeptides, useful as vaccine components for treating
or preventing streptococcal infections such as otitis media,
meningitis, and bacteraemia

Example 1; Page -: 113pp; English.

The invention describes an isolated polypeptide (I) with 70-90%
identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
comprising (I) is useful for therapeutic or prophylactic treatment of

CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardiae or
 CC Streptococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence represents a chimeric
 CC gene created from fragments and variant fragments of Streptococcus
 CC pneumoniae genes, described in the method of the invention.
 CC Note: This sequence does not appear in the specification but has
 CC been created according to information given in the invention.

CC Sequence 1378 AA;

Query Match 3.4%; Score 27; DB 23; Length 1378;
 Best Local Similarity 100.0%; Pred. No. 1.4e-16;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 326 RIARIPLRYRSNMWVPDSRPEQSPQ 352

Db 63 RIARIPLRYRSNMWVPDSRPEQSPQ 89

RESULT 78

AAV91938 standard; Protein; 168 AA.

AAV91938;

19-JUL-2000 (first entry)

S. pneumoniae 20 kDa human C3-degrading protein.

Human C3-degrading protein; 20 kDa; immunostimulatory; vaccine;
 inhibitor; inflammation; organ rejection; xenotransplantation.

Streptococcus pneumoniae.

WO200017370-A1.

30-MAR-2000.

24-SEP-1999; 99WO-US22362.

24-SEP-1998; 98US-0101736.

31-MAR-1999; 99US-0283094.

(MINU) UNIV MINNESOTA.

(AMCY) AMERICAN CYANAMID CO.

Hoseltter MK, Finkel DJ, Cheng Q, Green BA, Masi AW;

WPI; 2000-283594/24.

N-PSDB; AAA08556.

Isolated polypeptide is used to stimulate immune system and immunize or
 treat a mammalian subject against Streptococcus pneumoniae infection or
 colonization

Claim 8; Page 54-55; 63pp; English.

The present sequence, isolated from Streptococcus pneumoniae, is a
 human C3-degrading protein of about 20 kDa. This sequence may be part
 of a larger, approximately 92 kDa protein (see AAV91939), also having
 human C3-degrading activity. The DNA sequences (AAA08556-57) can be
 used for producing an immune response to Streptococcus pneumoniae in a

CC mammal. Antibodies against the proteins can be used to inhibit
 CC S. pneumoniae-mediated C3 degradation. C3-mediated inflammation and
 CC rejection in xenotransplantation can be inhibited by expressing the
 CC nucleic acid sequences on the surface of an organ of an animal. In
 CC particular, the polypeptides are useful for stimulating the immune
 CC system and are effective to immunize or treat a mammalian subject
 CC against Streptococcus pneumoniae infection or colonization.

CC Sequence 168 AA;

Query Match 3.1%; Score 25; DB 21; Length 168;
 Best Local Similarity 100.0%; Pred. No. 1.8e-15;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 328 ARIRPLRYRSNMWVPDSRPEQSPQ 352

Db 62 ARIRPLRYRSNMWVPDSRPEQSPQ 86

RESULT 79

AAAB12738 standard; Protein; 226 AA.

AAAB12738;

21-NOV-2000 (first entry)

Streptococcus pneumoniae NEM8 protein antigen SEQ ID NO: 71.

Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 otitis media; pneumonia; immunisation; bactericidal.

Streptococcus pneumoniae.

WO200039289-A2.

06-JUL-2000.

20-DEC-1999; 99WO-CA01218.

23-DEC-1998; 98US-0113800.

(BIOC-) BIOCHEM PHARMA INC.

Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

WPI; 2000-452397/39.

Streptococcal antigens useful for vaccinating against e.g. meningitis,
 otitis media, bacteraemia and/or pneumonia -

Claim 18; Fig 36; 106pp; English.

The present invention describes nucleic acids (I) encoding protein
 antigens (II) from Streptococcus pneumoniae. The protein antigens
 have bactericidal activity. The nucleic acids, encoding the proteins
 antigens, may be used for the recombinant production of the proteins
 they encode. The protein antigens may then be used as vaccines for the
 prevention and treatment of Streptococcal infections in mammals
 (especially humans) which result in, e.g. meningitis, otitis media,
 bacteraemia and/or pneumonia. The present sequence represents the
 S. pneumoniae NEM8 protein antigen.

CC Sequence 226 AA;

Query Match 3.0%; Score 24; DB 21; Length 226;
 Best Local Similarity 100.0%; Pred. No. 2.1e-14;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 266 DIDSLLKQLYKLPLSQHVESDGL 289

Db 1 DIDSLLKQLYKLPLSQHVESDGL 24


```
PD 27-DEC-2001.
XX
XX 19-JUN-2001; 2001WO-CA00908.
PF
XX 20-JUN-2000; 2000US-212683P.
PR
XX (SHIR-) SHIRE BIOCHEM INC.
XX PA
XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
PI WPI; 2002-122272/16.
XX
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PR or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia
XX
XX Example 1; Page -: 113pp; English.
PS
XX The invention describes an isolated polypeptide (I) with 70-90%
XX identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
XX BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
XX comprising (I) is useful for therapeutic or prophylactic treatment of
XX meningitis, otitis media, bacteraemia or pneumonia infection in an
XX individual susceptible to these disorders. (II) is also useful for
XX therapeutic or prophylactic treatment of any streptococcal bacterial
XX infection (e.g., caused by Streptococcus pneumoniae, group A
XX Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
XX as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novacida or
XX Streptococcus aureus) in an individual susceptible to the infection.
XX A polynucleotide (III) encoding (I) is useful in DNA immunisation
XX techniques. The Streptococcus polypeptides are useful in a diagnostic
XX test for S. pneumoniae infection. (III) is useful for designing DNA
XX probes for use in detecting the presence of Streptococcus in a biological
XX sample suspected of containing the bacteria. The DNA probes may also be
XX used for detecting circulating S. pneumonia nucleic acid in a sample for
XX diagnosing streptococcal infections. This sequence represents a truncate
XX of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
XX described in the method of the invention.
XX Note: This sequence does not appear in the specification but has
XX been created according to information given in the invention.
XX
XX SQ Sequence 428 AA;
XX
XX Query Match 3.0%; Score 24; DB 23; Length 428;
XX Best Local Similarity 100.0%; Pred. No. 3.8e-14;
XX Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 266 DIDSLLKQLYKLPISQRHVESDGL 289
XX |||||
XX 1 DIDSLLKQLYKLPISQRHVESDGL 24
XX
XX Db
XX
XX RESULT 83
XX AAB12734
XX ID AAB12734 standard; Protein; 555 AA.
XX
XX AC AAB12734;
XX
XX XX 21-NOV-2000 (first entry)
XX
XX DE Streptococcus pneumoniae NEW4 protein antigen SEQ ID NO:67.
XX
XX KM Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
XX KM prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX KM otitis media; pneumonia; immunisation; bactericidal.
XX
XX OS Streptococcus pneumoniae.
XX
XX PN WO200039299-A2.
XX
XX PD 06-JUL-2000.
XX
```

```
PF 20-DEC-1999; 99WO-CA01218.
XX
XX 23-DEC-1998; 98US-0113800.
PR
XX (BIOC-) BIOCHEM PHARMA INC.
XX PA
XX Hamel J, Brodeur BR, pineau I, Martin D, Rioux C, Charland N;
PI WPI; 2000-452397/39.
XX
XX DR Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX PT otitis media, bacteraemia and/or pneumonia
XX
XX PS Claim 18; Fig 32; 106pp; English.
XX
XX CC The present invention describes nucleic acids (I) encoding protein
XX antigens (II) from Streptococcus pneumoniae. The protein antigens
XX have bactericidal activity. The nucleic acids, encoding the protein
XX antigens, may be used for the recombinant production of the proteins
XX CC they encode. The protein antigens may then be used as vaccines for the
XX CC prevention and treatment of Streptococcal infections in mammals
XX (especially humans) which result in, e.g. meningitis, otitis media,
XX CC bacteraemia and/or pneumonia. The present sequence represents the
XX CC S. pneumoniae NEW4 protein antigen.
XX
XX SQ Sequence 555 AA;
XX
XX Query Match 3.0%; Score 24; DB 21; Length 555;
XX Best Local Similarity 100.0%; Pred. No. 4.8e-14;
XX Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 266 DIDSLLKQLYKLPISQRHVESDGL 289
XX |||||
XX 1 DIDSLLKQLYKLPISQRHVESDGL 24
XX
XX Db
XX
XX RESULT 84
XX AAU84047
XX ID AAU84047 standard; Peptide; 555 AA.
XX
XX AC AAU84047;
XX
XX XX 08-MAY-2002 (first entry)
XX
XX DE Truncated variant of S. pneumoniae BVH-11, NEW4.
XX
XX KM BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
XX KM pneumonia; streptococcal bacterial infection; mutant; mutein.
XX
XX OS Streptococcus pneumoniae.
XX
XX PN WO200198334-A2.
XX
XX PD 27-DEC-2001.
XX
XX PF 19-JUN-2001; 2001WO-CA00908.
XX
XX PR 20-JUN-2000; 2000US-212683P.
XX
XX PA (SHIR-) SHIRE BIOCHEM INC.
XX
XX PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX
XX DR WPI; 2002-122272/16.
XX
XX PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX PT epitope-bearing polypeptides, useful as vaccine components for treating
XX PT or preventing streptococcal infections such as otitis media,
XX PT meningitis, and bacteraemia
XX
XX PS Example 1; Page -: 113pp; English.
XX
```

CC The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novardi or
CC Streptococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a truncate
CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
CC described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.

SQ Sequence 555 AA;

Query Match 3.0%; Score 24; DB 23; Length 555;
Best Local Similarity 100.0%; Pred. No. 4.8e-14;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 266 DIDSLLKQLYKPLSQRHVESDGL 289
DB 1 DIDSLLKQLYKPLSQRHVESDGL 24

RESULT 85

AAU84051 standard; Peptide; 999 AA.

AC AAU84051;

DT 08-MAY-2002 (first entry)

DE S. pneumoniae derived chimeric peptide, NEW17.

XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;

KM pneumonia; streptococcal bacterial infection; mutant; mutein;

KM BVH-11-2.

XX Streptococcus pneumoniae.

OS Synthetic.

XX WO200198334-A2.

PN 27-DEC-2001.

PF 19-JUN-2001; 2001WO-CA00908.

PR 20-JUN-2000; 2000US-212683P.

XX (SHIR-) SHIRE BIOCHEM INC.

PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

DR WPI; 2002-122272/16.

XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX PT epitope-bearing polypeptides, useful as vaccine components for treating
XX PT or preventing streptococcal infections such as otitis media,
XX PT meningitis, and bacteraemia

PS Example 1; Page -; 113pp; English.

XX The invention describes an isolated polypeptide (I) with 70-90%
CC

CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novardi or
CC Streptococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a chimeric
CC gene created from fragments and variant fragments of Streptococcus
CC pneumoniae genes, described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.

SQ Sequence 999 AA;

Query Match 3.0%; Score 24; DB 23; Length 999;
Best Local Similarity 100.0%; Pred. No. 8.2e-14;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 266 DIDSLLKQLYKPLSQRHVESDGL 289
DB 2 DIDSLLKQLYKPLSQRHVESDGL 25

RESULT 86

AAU84052 standard; Peptide; 999 AA.

AC AAU84052;

DT 08-MAY-2002 (first entry)

DE S. pneumoniae derived chimeric peptide, NEW20.

XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;

KM pneumonia; streptococcal bacterial infection; mutant; mutein;

KM BVH-11-2.

XX Streptococcus pneumoniae.

OS Synthetic.

XX WO200198334-A2.

PN 27-DEC-2001.

PF 19-JUN-2001; 2001WO-CA00908.

PR 20-JUN-2000; 2000US-212683P.

XX (SHIR-) SHIRE BIOCHEM INC.

PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

DR WPI; 2002-122272/16.

XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX PT epitope-bearing polypeptides, useful as vaccine components for treating
XX PT or preventing streptococcal infections such as otitis media,
XX PT meningitis, and bacteraemia

PS Example 1; Page -; 113pp; English.

XX The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC

CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novae or
CC Streptococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a chimeric
CC gene created from fragments and variant fragments of Streptococcus
CC pneumoniae genes, described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.

SQ Sequence 999 AA;

Query Match 3.0%; Score 24; DB 23; Length 999;
Best Local Similarity 100.0%; Pred. No. 8.2e-14;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 266 DIDLKQLKYLPLSQRHVESDGL 289
Db 572 DIDLKQLKYLPLSQRHVESDGL 595

RESULT 87

ID AAU84058 standard; Peptide; 1126 AA.

AC AAU84058;

DT 08-MAY-2002 (first entry)

DE S. pneumoniae derived chimeric peptide, NEW31.

KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutein;
KW BVH-11-2.

OS Streptococcus pneumoniae.

OS Synthetic.

PN WO200198334-A2.

PD 27-DEC-2001.

PF 19-JUN-2001; 2001WO-CA00908.

PR 20-JUN-2000; 2000US-212683P.

PA (SHIR-) SHIRE BIOCHEM INC.

PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

DR WPI; 2002-122272/16.

PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia

PS Example 1; Page -; 113pp; English.

CC The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)

CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novae or
CC Streptococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a chimeric
CC gene created from fragments and variant fragments of Streptococcus
CC pneumoniae genes, described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.

SQ Sequence 1126 AA;

Query Match 3.0%; Score 24; DB 23; Length 1126;
Best Local Similarity 100.0%; Pred. No. 9.1e-14;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 266 DIDLKQLKYLPLSQRHVESDGL 289
Db 2 DIDLKQLKYLPLSQRHVESDGL 25

RESULT 88

ID AAU84056 standard; Peptide; 1238 AA.

AC AAU84056;

DT 08-MAY-2002 (first entry)

DE S. pneumoniae derived chimeric peptide, NEW29.

KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutein;
KW BVH-11-2.

OS Streptococcus pneumoniae.

OS Synthetic.

PN WO200198334-A2.

PD 27-DEC-2001.

PF 19-JUN-2001; 2001WO-CA00908.

PR 20-JUN-2000; 2000US-212683P.

PA (SHIR-) SHIRE BIOCHEM INC.

PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

DR WPI; 2002-122272/16.

PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia

PS Example 1; Page -; 113pp; English.

CC The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of

meningitis, otitis media, bacteraemia or pneumonia infection in an individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by Streptococcus pneumoniae, group A Streptococcus such as Streptococcus pyogenes, group B Streptococcus such as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. norcardia or Streptococcus aureus) in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The Streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (III) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating S. pneumonia nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a chimeric gene created from fragments and variant fragments of Streptococcus pneumoniae genes, described in the method of the invention. Note: This sequence does not appear in the specification but has been created according to information given in the invention.

Sequence 1238 AA;

Query Match 3.0%; Score 24; DB 23; Length 1238;
Best Local Similarity 100.0%; Pred. No. 9.9e-14;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 266 DIDSLKQLYKPLPSQRHVESDGL 289
|||

Db 2 DIDSLKQLYKPLPSQRHVESDGL 25

RESULT 89

AAU84057
ID AAU84057 standard; Peptide; 1365 AA.

AAU84057;

08-MAY-2002 (first entry)

S. pneumoniae derived chimeric peptide, NEM30.

BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
pneumonia; streptococcal bacterial infection; mutant; mutein;

BVH-11-2.

Streptococcus pneumoniae.
Synthetic.

WO200198334-A2.

27-DEC-2001.

19-JUN-2001; 2001WO-CA00908.

20-JUN-2000; 2000US-212683P.

(SHIR-) SHIRE BIOCHEM INC.

Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

WPI; 2002-122272/16.

New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
epitope-bearing polypeptides, useful as vaccine components for treating

or preventing streptococcal infections such as otitis media,
meningitis, and bacteraemia

Example 1; Page -; 113pp; English.

The invention describes an isolated polypeptide (I) with 70-90%

identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)

comprising (I) is useful for therapeutic or prophylactic treatment of
meningitis, otitis media, bacteraemia or pneumonia infection in an

individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by Streptococcus pneumoniae, group A Streptococcus such as Streptococcus pyogenes, group B Streptococcus such as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. norcardia or Streptococcus aureus) in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The Streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (III) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating S. pneumonia nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a chimeric gene created from fragments and variant fragments of Streptococcus pneumoniae genes, described in the method of the invention. Note: This sequence does not appear in the specification but has been created according to information given in the invention.

Sequence 1365 AA;

Query Match 3.0%; Score 24; DB 23; Length 1365;
Best Local Similarity 100.0%; Pred. No. 1.1e-13;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 266 DIDSLKQLYKPLPSQRHVESDGL 289
|||

Db 2 DIDSLKQLYKPLPSQRHVESDGL 25

RESULT 90

AAU83827
ID AAU83827 standard; Peptide; 94 AA.

AAU83827;

08-MAY-2002 (first entry)

S. pneumoniae antigenic peptide 404.9.

BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
pneumonia; streptococcal bacterial infection.

Streptococcus pneumoniae.

WO200198334-A2.

27-DEC-2001.

19-JUN-2001; 2001WO-CA00908.

20-JUN-2000; 2000US-212683P.

(SHIR-) SHIRE BIOCHEM INC.

Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

WPI; 2002-122272/16.

New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
epitope-bearing polypeptides, useful as vaccine components for treating

or preventing streptococcal infections such as otitis media,
meningitis, and bacteraemia

Example 3; Figure 11; 113pp; English.

The invention describes an isolated polypeptide (I) with 70-90%

identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)

comprising (I) is useful for therapeutic or prophylactic treatment of
meningitis, otitis media, bacteraemia or pneumonia infection in an

individual susceptible to these disorders. (II) is also useful for
therapeutic or prophylactic treatment of any streptococcal bacterial

infection (e.g., caused by Streptococcus pneumoniae, group A

CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. norcardia or
 CC Streptococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (iii) encoding (i) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (iii) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence represents an
 CC antigenic peptide derived from the S. pneumoniae genome, described in the
 CC method of the invention.

CC Sequence 94 AA;

Query Match 2.9%; Score 23; DB 23; Length 94;
 Best Local Similarity 100.0%; Pred. No. 8.8e-14;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 DGGYVTSHGHDHYHNGKVPYDA 77
 DB 1 DGGYVTSHGHDHYHNGKVPYDA 23

RESULT 91
 ID AAY05752 standard; Protein; 163 AA.

AC AAY05752;
 DT 19-JUL-1999 (first entry)

DE Streptococcus pneumoniae 20 kDa complement C3-degrading protease.

KM Human complement C3-degrading protease; vaccine; infection;
 KM meningitis; pneumonia; xerobransplantation; transplant rejection;
 KM inflammation.

OS Streptococcus pneumoniae.

Key Location/Qualifiers
 FT 1..58
 FT /note="Claimed protein of Claim 12"
 FT 90..132
 FT Protein /note="Claimed protein of Claim 13"

W09915675-A1.

01-APR-1999.

24-SEP-1998; 98WO-US20186.

24-SEP-1997; 97US-0059907.

(AMCY) AMERICAN CYANAMID CO.

(MINU) UNIV MINNESOTA.

Cheng Q, Finkel DJ, Green BA, Hostetter MK, Masi AM;

WPI, 1999-254719/21.

N-PSDB; AAX25393.

New isolated human complement C3-degrading proteinase

Claim 7, Page 50; 66pp; English.

The present sequence is a 20 kDa protein of Streptococcus pneumoniae
 serotype 3 that is capable of degrading human complement protein C3
 (HCP3). The sequence was deduced from the open reading frame of
 an isolated gene (see AAX25393). Alignment of this gene sequence
 with the S. pneumoniae (serotype 4) genome showed that the open
 reading frame that codes for the 20 kDa protein may be part of a
 larger open reading frame, i.e. a 2163 bp (see AAX25394) in the

CC serotype 4 genome, which encodes a protein of approximately 79
 CC kDa (see AAY05753). Amino acids 1-58 and 90-132 of the 20 kDa
 CC protein have substantial sequence identity with amino acids 170-227
 CC and 258-300 of the 79 kDa protein. Proteins and peptides or
 CC polypeptides containing these regions, and DNA sequences encoding
 CC them, are claimed. HCP3 proteases and polypeptides can be used as
 CC immune system stimulating compositions (claimed). They can produce
 CC an immune response against S. pneumoniae to immunize or treat a
 CC mammalian subject against infection or colonization (claimed).
 CC They can produce a B cell response, a T cell response, an
 CC epithelial cell response, or an endothelial cell response
 CC (claimed). The expression of the proteins on the surface of an
 CC organ of an animal used in xenotransplantation can be used to
 CC inhibit C3-mediated inflammation and rejection.

CC Sequence 163 AA;

Query Match 2.1%; Score 17; DB 20; Length 163;
 Best Local Similarity 100.0%; Pred. No. 9e-08;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 303 GVAVPHGDHYHFRIPYSQ 319
 DB 37 GVAVPHGDHYHFRIPYSQ 53

RESULT 92
 ID AAY91286 standard; Protein; 381 AA.

AC AAY91286;

DT 30-MAY-2000 (first entry)

DE Group B Streptococcus protein sequence SEQ ID NO:12.

KM Group B Streptococcus; Streptococcus agalactiae; protein antigen;
 KM vaccine; screening; immunogen; detection; diagnosis; infection;
 KM antibody; antibody; antibacterial.

OS Streptococcus agalactiae.

W0200006736-A2.

10-FEB-2000.

27-JUL-1999; 99WO-GB02444.

27-JUL-1998; 98GB-0016335.

19-MAR-1999; 99US-0125163.

(MICR-) MICROBIAL TECHNICS LTD.

Le Page RWF, Wells JM, Hamillfy SB;

WPI, 2000-195299/17.

New Group B Streptococcus protein, useful as vaccine, for diagnosis of
 PT Streptococcal infections and for screening of antibodies or antibodies

Claim 1; Fig 1; 123pp; English.

AA05803 to AA05872 encode proteins, polypeptides and peptides (given
 CC in AAY91275 to AAY91343) isolated from Group B Streptococcus (GBS), also
 CC known as Streptococcus agalactiae. The GBS polynucleotides and
 CC polypeptides have antibacterial activity. Immunogenic compositions
 CC comprising GBS polynucleotides or polypeptides can be used as vaccines
 CC and for the treatment or prophylaxis of GBS infection. The
 CC polynucleotides and polypeptides can also be used in the detection of GBS
 CC and for screening DNA encoding bacterial cell envelope associated or
 CC secreted antigens in gram positive bacteria. AA05873 to AA05941
 CC represent primers used in the exemplification of the present invention.

SQ Sequence 381 AA;

Query Match 2.1%; Score 17; DB 21; Length 381;

Best Local Similarity 100.0%; Pred. No. 2e-07;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 69 YNGKVPYDAIISELLM 85
 |||||
 DB 92 YNGKVPYDAIISELLM 108

RESULT 93

AAV27347

ID AAV27347 standard; protein; 793 AA.

XX AAV27347;

DT 15-NOV-1999 (first entry)

DE Group B Streptococcus (GBS) antigen (clone 3).

KM Group B Streptococcus; GBS; antigen; vaccine; Streptococcus infection;

KW sepsis; meningitis; pneumonia; immunocompromise; diabetes; liver disease;

XX cancer; velinary; mastitis.

OS Streptococcus sp.

PN WO9942588-A2.

PD 26-AUG-1999.

PF 17-FEB-1999; 99WO-CA00114.

PR 20-FEB-1998; 98US-0075425.

(BIOC-) BIOCHEM VACCINS INC.

PI Boyer M, Brodeur BR, Charlebois I, Hamel J, Martin D;

XX Riaux C;

DR WPI; 1999-540309/45.

XX N-PSDB; AAX91105.

PT Novel group B Streptococcus antigens - useful as vaccine

XX compositions for prophylaxis or therapy of Streptococcus infections

XX Claim 26; Fig 3C; 154pp; English.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

RESULT 94

AAV91283

ID AAV91283 standard; Protein; 822 AA.

XX AAV91283;

DT 30-MAY-2000 (first entry)

DE Group B Streptococcus protein sequence SEQ ID NO:9.

KM Group B Streptococcus; Streptococcus agalactiae; protein antigen;

KW vaccine; screening; immunogen; detection; diagnosis; infection;

XX antibody; affibody; antibacterial.

OS Streptococcus agalactiae.

PN WO200006736-A2.

PD 10-FEB-2000.

PF 27-JUL-1999; 99WO-GB02444.

PR 27-JUL-1998; 98GB-0016335.

PR 19-MAR-1999; 99US-0125163.

(MICR-) MICROBIAL TECHNICS LTD.

PI Le Page RWF, Wells JM, Hanniffy SB;

XX WPI; 2000-195299/17.

PT New Group B Streptococcus protein, useful as vaccine, for diagnosis of

XX Streptococcal infections and for screening of antibodies or affibodies

XX Claim 1; Fig 1; 123pp; English.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

RESULT 95

AAU00028

ID AAU00028 standard; Protein; 822 AA.

XX AAU00028;

DT 11-MAY-2001 (first entry)

DE Streptococcus agalactiae GBS36#1.

KM Group B streptococci protein 36; GBS36; immunogen; vaccine;

KW antibody; necrotizing fasciitis; scarlet fever; sepsis; impetigo;

XX bacterial meningitis; otitis media; community-acquired pneumonia.

OS Streptococcus agalactiae.

PN WO200006736-A2.

PD 10-FEB-2000.

PF 27-JUL-1999; 99WO-GB02444.

PR 27-JUL-1998; 98GB-0016335.

PR 19-MAR-1999; 99US-0125163.

(MICR-) MICROBIAL TECHNICS LTD.

PI Le Page RWF, Wells JM, Hanniffy SB;

XX WPI; 2000-195299/17.

PT New Group B Streptococcus protein, useful as vaccine, for diagnosis of

XX Streptococcal infections and for screening of antibodies or affibodies

XX Claim 1; Fig 1; 123pp; English.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

OS	Streptococcus agalactiae.
XX	
FH	Key
FT	Location/Qualifiers
FT	Region
FT	/label= Histidine triad motif
FT	/note= "Proposed to be involved in metal binding"
FT	219..224
FT	/label= Histidine triad motif
FT	/note= "Proposed to be involved in metal binding"
FT	353..358
FT	/label= Histidine triad motif
FT	/note= "Proposed to be involved in metal binding"
FT	444..449
FT	/label= Histidine triad motif
FT	/note= "Proposed to be involved in metal binding"
FT	584..589
FT	/label= Histidine triad motif
FT	/note= "Proposed to be involved in metal binding"
XX	
PN	WO200114421-A1.
PD	
XX	01-MAR-2001.
PF	
XX	25-AUG-2000; 2000WO-US23417.
XX	
PR	25-AUG-1999; 99US-0150750.
XX	
PA	(MED1-) MEDIMMUNE INC.
XX	
PI	Koenig S, Heinrichs J, Johnson LS, Adamou JE;
XX	
XX	WPI; 2001-211305/21.
DR	N-PSDB; AAS00038.
XX	
PT	New polypeptides obtained from group A or B streptococci, especially
PT	Streptococcus aureus homologous to Sp36 protein of Streptococcus
PT	pneumoniae useful as antibacterial vaccines -
XX	
PS	Claim 4; Fig 2; 62pp; English.
XX	
CC	The sequence represents Streptococcus agalactiae Group B Streptococci
CC	protein 36, GBS36. A recombinant cell producing GBS36, GBS36(2) or
CC	GBS36 is useful as a vaccine for vaccinating an animal, preferably a
CC	human against infection by a bacterial organism such as a streptococcal
CC	or staphylococcal bacteria, and for treating a disease caused by group A
CC	streptococci, group B streptococci or Staphylococcus aureus in an animal
CC	preferably a human. Vaccines and antibodies against the proteins of the
CC	invention are useful in prophylaxis and/or treatment of diseases such as
CC	necrotising fasciitis, scarlet fever, sepsis, impetigo, bacterial
CC	meningitis, otitis media, community-acquired pneumonia and many diseases
CC	of newborns. The proteins are also used as immunogens to stimulate the
CC	production of antibodies for use in passive immunotherapy, for use as
CC	diagnostic reagents and for use as reagents in other processes such as
CC	affinity chromatography.
XX	
SQ	Sequence 822 AA:
QY	
Query Match	2.1%; Score 17; DB 22; Length 822;
Best Local Similarity	100.0%; Pred. No. 3.9e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
Db	69 YNGKVPYDAIISELLM 85
	92 YNGKVPYDAIISELLM 108
RESULT 96	
AAU00030	AAU00030 standard; Protein; 822 AA.
XX	
AC	AAU00030;
XX	
XX	11-MAY-2001 (first entry)

XX	Streptococcus agalactiae GBS36#2.
DE	
XX	Group B streptococci protein 36; GBS36; immunogen; vaccine;
KW	antibody; necrotising fasciitis; scarlet fever; sepsis; impetigo;
KW	bacterial meningitis; otitis media; community-acquired pneumonia.
XX	
OS	Streptococcus agalactiae.
XX	
FH	Key
FT	Misc_feature
FT	Location/Qualifiers
FT	1 /note= "Encoded by GNG using prokaryote alternative genetic code"
FT	85..90
Region	/label= Histidine triad_motif
FT	/note= "Proposed to be involved in metal binding"
FT	219..224
Region	/label= Histidine triad_motif
FT	/note= "Proposed to be involved in metal binding"
FT	353..358
Region	/label= Histidine triad_motif
FT	/note= "Proposed to be involved in metal binding"
FT	444..449
Region	/label= Histidine triad_motif
FT	/note= "Proposed to be involved in metal binding"
FT	584..589
Region	/label= Histidine triad_motif
FT	/note= "Proposed to be involved in metal binding"
XX	
PV	WO200114421-A1.
XX	
PD	01-MAR-2001.
XX	
PF	25-AUG-2000; 2000WO-US23417.
XX	
PR	25-AUG-1999; 99US-0150750.
XX	
PA	(MED1-) MEDIMMUNE INC.
PI	Koenig S, Heinrichs J, Johnson LS, Adamou JE;
DR	WPI; 2001-211305/21.
XX	N-PADB; AAS00038.
PT	New polypeptides obtained from group A or B streptococci , especially
PT	Staphylococcus aureus homologous to Sp36 protein of Streptococcus
PT	pneumoniae useful as antibacterial vaccines -
XX	
PS	Example 2; Fig 5c; 62pp; English.
XX	
CC	The sequence represents Streptococcus agalactiae Group B Streptococci
CC	protein 36, GBS36. A recombinant cell producing GAS36, GAS36(2) or
CC	GBS36 is useful as a vaccine for vaccinating an animal, preferably a
CC	human against infection by a bacterial organism such as a streptococcal
CC	human against infection by a bacterial organism caused by group A
CC	streptococci, group B streptococci or Staphylococcus aureus in an animal
CC	preferably a human. Vaccines and antibodies against the proteins of the
CC	invention are useful in prophylaxis and/or treatment of diseases such as
CC	necrotising fasciitis, scarlet fever, sepsis, impetigo, bacterial
CC	meningitis, otitis media, community-acquired pneumonia and many diseases
CC	of newborns. The proteins are also used as immunogens to stimulate the
CC	production of antibodies for use in passive immunotherapy, for use as
CC	diagnostic reagents and for use as reagents in other processes such as
CC	affinity chromatography.
XX	
SQ	Sequence 822 AA:
Query Match	2.1%; Score 17; DB 22; Length 822;
Best Local Similarity	100.0%; Pred. No. 3.9e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
69 YNKGVPYDATISELLM 85	

Db 92 YNGKVPYDAIISELLM 108

RESULT 97

ID ABP28904 standard; Protein; 822 AA.

AC ABP28904;

DT 02-JUL-2002 (first entry)

DE Streptococcus polypeptide SEQ ID NO 6984.

KM Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;

KW group A streptococcus; Streptococcus pyogenes; antibacterial;

XX antiinflammatory; infection; vaccine; meningitis; gene therapy.

OS Streptococcus agalactiae.

PN WO200234771-A2.

PD 02-MAY-2002.

PF 29-OCT-2001; 2001WO-GB04789.

PR 27-OCT-2000; 2000GB-0026333.

PR 24-NOV-2000; 2000GB-0028727.

PR 07-MAR-2001; 2001GB-0005640.

PA (CHIR-) CHIRON SPA.

PA (GENO-) INST GENOMIC RES.

PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;

PI Tettelein H;

DR WPI; 2002-352536/38.

DR N-PSDB; ABN69535.

PT New Streptococcus protein for the treatment or prevention of infection

PT or disease caused by Streptococcus bacteria, such as meningitis, and

PT for detecting a compound that binds to the protein -

PS Claim 1; Page 3861; 4525pp; English.

CC The invention relates to a protein (ABP25413-ABP30895) from group B

CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS

CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in

CC the specification. The proteins have antibacterial and antiinflammatory

CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and

CC antibodies that bind (I) are used in the manufacture of medicaments for

CC the treatment or prevention of infection or disease caused by

CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.

CC Nucleic acids encoding (I) are used to detect Streptococcus in a

CC biological sample. (I) is used to determine whether a compound binds to

CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be

CC used as a vaccine or diagnostic composition. The disease caused by

CC Streptococcus that is prevented or treated may be meningitis. Nucleic

CC acid encoding (I) may be used to recombinantly produce (I) and may be

CC used in gene therapy. Antibodies to (I) are used for affinity

CC chromatography, immunoassays, and distinguishing/identifying

CC Streptococcus proteins.

SQ Sequence 822 AA;

QY Query Match 2.1%; Score 17; DB 23; Length 822;

DB Best Local Similarity 100.0%; Pred. No. 3.9e-07;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

69 YNGKVPYDAIISELLM 85

92 YNGKVPYDAIISELLM 108

RESULT 98

ID ABP29703 standard; Protein; 822 AA.

AC ABP29703;

DT 02-JUL-2002 (first entry)

DE Streptococcus polypeptide SEQ ID NO 8582.

KM Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;

KW group A streptococcus; Streptococcus pyogenes; antibacterial;

XX antiinflammatory; infection; vaccine; meningitis; gene therapy.

OS Streptococcus agalactiae.

PN WO200234771-A2.

PD 02-MAY-2002.

PF 29-OCT-2001; 2001WO-GB04789.

PR 27-OCT-2000; 2000GB-0026333.

PR 24-NOV-2000; 2000GB-0028727.

PR 07-MAR-2001; 2001GB-0005640.

PA (CHIR-) CHIRON SPA.

PA (GENO-) INST GENOMIC RES.

PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;

PI Tettelein H;

DR WPI; 2002-352536/38.

DR N-PSDB; ABN70334.

PT New Streptococcus protein for the treatment or prevention of infection

PT or disease caused by Streptococcus bacteria, such as meningitis, and

PT for detecting a compound that binds to the protein -

PS Claim 1; Page 3969; 4525pp; English.

CC The invention relates to a protein (ABP25413-ABP30895) from group B

CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS

CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in

CC the specification. The proteins have antibacterial and antiinflammatory

CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and

CC antibodies that bind (I) are used in the manufacture of medicaments for

CC the treatment or prevention of infection or disease caused by

CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.

CC Nucleic acids encoding (I) are used to detect Streptococcus in a

CC biological sample. (I) is used to determine whether a compound binds to

CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be

CC used as a vaccine or diagnostic composition. The disease caused by

CC Streptococcus that is prevented or treated may be meningitis. Nucleic

CC acid encoding (I) may be used to recombinantly produce (I) and may be

CC used in gene therapy. Antibodies to (I) are used for affinity

CC chromatography, immunoassays, and distinguishing/identifying

CC Streptococcus proteins.

SQ Sequence 822 AA;

QY Query Match 2.1%; Score 17; DB 23; Length 822;

DB Best Local Similarity 100.0%; Pred. No. 3.9e-07;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

69 YNGKVPYDAIISELLM 85

92 YNGKVPYDAIISELLM 108

RESULT 99

ID AAB12746 standard; Protein; 823 AA.

```

XX AAB12746;
AC 21-NOV-2000 (first entry)
XX
XX
XX
XX
XX Streptococcus pneumoniae GBS BVH-71 protein antigen SEQ ID NO:81.
XX
XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
XX prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX otitis media; pneumonia; immunisation; bactericidal.
XX
XX Streptococcus pneumoniae.
XX
XX WO200039299-A2.
XX
XX 06-JUL-2000.
XX
XX 20-DEC-1999; 99WO-CA01218.
XX
XX 23-DEC-1998; 98US-0113800.
XX
XX (BIOC-) BIOCHEM PHARMA INC.
XX
XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX
XX WPI: 2000-452397/39.
XX
XX N-PSDB; AAA65740.
XX
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX otitis media, bacteraemia and/or pneumonia -
XX
XX Claim 18; Fig 46; 106pp; English.
XX
XX The present invention describes nucleic acids (I) encoding protein
XX antigens (II) from Streptococcus pneumoniae. The protein antigens
XX have bactericidal activity. The nucleic acids, encoding the protein
XX antigens, may be used for the recombinant production of the proteins
XX they encode. The protein antigens may then be used as vaccines for the
XX prevention and treatment of Streptococcal infections in mammals
XX (especially humans) which result in, e.g. meningitis, otitis media,
XX bacteraemia and/or pneumonia. The present sequence represents the
XX S. pneumoniae GBS BVH-71 protein antigen.
XX
XX Sequence 823 AA:
XX
XX Query Match 2.1%; Score 17; DB 21; Length 823;
XX Best Local Similarity 100.0%; Pred. No. 3.9e-07;
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 69 YNGKVPYDAIISELLM 85
XX |||||
XX DB 92 YNGKVPYDAIISELLM 108
XX
XX RESULT 100
XX AAB12747
XX ID AAB12747 standard; Protein; 824 AA.
XX
XX AAB12747;
XX
XX 21-NOV-2000 (first entry)
XX
XX Streptococcus pneumoniae GBS BVH-71 protein antigen SEQ ID NO:83.
XX
XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
XX prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX otitis media; pneumonia; immunisation; bactericidal.
XX
XX Streptococcus pneumoniae.
XX
XX WO200039299-A2.
XX
XX 06-JUL-2000.
XX

```

```

XX
XX 20-DEC-1999; 99WO-CA01218.
XX
XX 23-DEC-1998; 98US-0113800.
XX
XX (BIOC-) BIOCHEM PHARMA INC.
XX
XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX
XX WPI: 2000-452397/39.
XX
XX N-PSDB; AAA65741.
XX
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX otitis media, bacteraemia and/or pneumonia -
XX
XX Claim 18; Fig 46; 106pp; English.
XX
XX The present invention describes nucleic acids (I) encoding protein
XX antigens (II) from Streptococcus pneumoniae. The protein antigens
XX have bactericidal activity. The nucleic acids, encoding the protein
XX antigens, may be used for the recombinant production of the proteins
XX they encode. The protein antigens may then be used as vaccines for the
XX prevention and treatment of Streptococcal infections in mammals
XX (especially humans) which result in, e.g. meningitis, otitis media,
XX bacteraemia and/or pneumonia. The present sequence represents the
XX S. pneumoniae GBS BVH-71 protein antigen.
XX
XX Sequence 824 AA:
XX
XX Query Match 2.1%; Score 17; DB 21; Length 824;
XX Best Local Similarity 100.0%; Pred. No. 3.9e-07;
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 69 YNGKVPYDAIISELLM 85
XX |||||
XX DB 92 YNGKVPYDAIISELLM 108
XX

```

Search completed: November 14, 2003, 10:58:42
Job time : 473 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 10:50:50 ; Search time 40 Seconds
(Without alignments)
1913.757 Million cell updates/sec

Title: US-09-765-271-56

Perfect score: 796
Sequence: 1 SYELGLYQARTVKNRRVSY.....KLIALIKGSPNSVSXKXIN 796

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 200 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	635	87.3	802	2	C95136	conserved domain p
2	616	77.4	828	2	E98004	hypothetical prote
3	107	13.4	855	2	D98004	histidine Motif-Co
4	60	7.5	819	2	B95136	conserved domain p
5	60	7.5	839	2	G95115	conserved domain p
6	60	7.5	853	2	C97985	hypothetical prote
7	32	4.0	1039	2	H95115	conserved domain p
8	32	4.0	1039	2	D97985	hypothetical prote
9	17	2.1	822	2	T46758	hypothetical prote
10	10	1.3	182	2	P97985	hypothetical prote
11	9	1.1	921	2	A48184	hypothetical prote
12	9	1.1	921	2	A45183	transcription init
13	8	1.0	71	2	E72853	TBP-associated fac
14	8	1.0	284	2	S58921	AcOrf-29 protein -
15	8	1.0	284	2	S58922	tropomyosin isofor
16	8	1.0	292	2	T49361	probable transport
17	8	1.0	294	2	T27791	hypothetical prote
18	8	1.0	448	2	S05355	hypothetical prote
19	8	1.0	484	2	F82303	probable multidrug
20	8	1.0	488	2	B89557	hypothetical prote
21	8	1.0	795	1	S73830	endopeptidase La (
22	8	1.0	825	2	T00818	hypothetical prote
23	8	1.0	840	2	C75313	probable proteinase
24	8	1.0	945	2	A64714	helicase - Helicob
25	8	1.0	1289	2	B70748	probable smc prote
26	8	1.0	1659	2	H97926	hypothetical prote
27	8	1.0	1659	2	G95057	endo-beta-N-acetyl
28	7	0.9	44	2	S55073	env polypeptide -
29	7	0.9	49	2	S55072	env polypeptide -

30	7	0.9	49	2	D70091	hypothetical prote
31	7	0.9	68	2	D87657	hypothetical prote
32	7	0.9	70	2	A75134	hypothetical prote
33	7	0.9	90	2	C90404	conserved domain p
34	7	0.9	93	2	C75396	hypothetical prote
35	7	0.9	94	2	E75556	hypothetical prote
36	7	0.9	100	1	BVMYBA	chaperonin groES -
37	7	0.9	100	1	BVMY7B	chaperonin groES -
38	7	0.9	101	2	T38824	thioltransferase -
39	7	0.9	101	2	E83874	ribosomal protein
40	7	0.9	106	2	I64171	hypothetical prote
41	7	0.9	110	2	D87498	HesB/YadR/Yhf fam
42	7	0.9	110	2	C86891	hypothetical prote
43	7	0.9	117	2	AG3598	transcription regu
44	7	0.9	121	2	T44473	conserved domain p
45	7	0.9	124	1	NRPRH	pancreatic ribonuc
46	7	0.9	128	2	H72500	hypothetical prote
47	7	0.9	129	2	T36220	mult-like protein
48	7	0.9	130	2	S76934	hypothetical prote
49	7	0.9	131	2	T35461	mult-like protein
50	7	0.9	136	2	B65240	hypothetical prote
51	7	0.9	136	2	D64964	hypothetical prote
52	7	0.9	136	2	T19240	hypothetical prote
53	7	0.9	139	2	B81286	probable membrane
54	7	0.9	143	2	A71717	hypothetical prote
55	7	0.9	145	2	AF1219	Salmonella enteric
56	7	0.9	147	2	D83261	probable transcrip
57	7	0.9	148	2	A86079	PTS system, fructo
58	7	0.9	148	2	B91232	PTS system, fructo
59	7	0.9	148	2	D48649	frvA protein - Esc
60	7	0.9	154	2	S39873	hypothetical prote
61	7	0.9	157	2	AC2947	ribonuclease H [lm
62	7	0.9	157	2	G98335	ribonuclease H PAI
63	7	0.9	157	2	AF3240	conserved domain p
64	7	0.9	158	1	D63256	hypothetical prote
65	7	0.9	158	2	A89943	transcription elon
66	7	0.9	165	2	C83415	hypothetical prote
67	7	0.9	166	2	A86450	probable glycine c
68	7	0.9	178	2	C81235	transcription anti
69	7	0.9	179	2	H71920	adenine phosphorib
70	7	0.9	181	2	A70178	conserved domain p
71	7	0.9	185	2	F75487	V-type ATP synthas
72	7	0.9	191	2	T35792	beta-glucosidase -
73	7	0.9	197	1	S01249	Cytochrome c-L pre
74	7	0.9	201	2	S06147	GTP-binding protei
75	7	0.9	203	2	F90658	hypothetical prote
76	7	0.9	203	2	F85509	hypothetical prote
77	7	0.9	210	2	S62443	diadenosine 5', 5'
78	7	0.9	211	2	AH0899	conserved domain p
79	7	0.9	213	2	T01627	probable ATP bindi
80	7	0.9	219	2	F83144	conserved domain p
81	7	0.9	219	2	B75533	hypothetical prote
82	7	0.9	226	2	D65105	hypothetical prote
83	7	0.9	226	2	A98133	hypothetical prote
84	7	0.9	226	2	D85978	hypothetical prote
85	7	0.9	226	2	G81894	probable integral
86	7	0.9	226	2	D81125	hypothetical prote
87	7	0.9	232	2	C83744	two-component resp
88	7	0.9	237	2	G87286	conserved domain p
89	7	0.9	238	2	I67638	proteasome activat
90	7	0.9	240	2	D75203	hypothetical prote
91	7	0.9	241	2	C71725	transcription init
92	7	0.9	241	2	H97723	hypothetical prote
93	7	0.9	255	2	T03482	iron(III) dicitrat
94	7	0.9	257	2	I36025	keratin-like prote
95	7	0.9	260	2	D86722	hypothetical prote
96	7	0.9	261	2	G70980	transcription init
97	7	0.9	261	2	D71037	hypothetical prote
98	7	0.9	269	2	T32639	hypothetical prote
99	7	0.9	271	1	ELK12	pancreatic elastas
100	7	0.9	279	2	S62364	L-2-chlorotriptoni
101	7	0.9	281	2	T43945	ribosomal protein
102	7	0.9	284	2	A44980	tropomyosin, obliq

103	7	0.9	284	2	A33085
104	7	0.9	284	2	S38381
105	7	0.9	284	2	D82486
106	7	0.9	289	2	T25831
107	7	0.9	293	2	S63356
108	7	0.9	295	2	B95298
109	7	0.9	296	2	T20005
110	7	0.9	298	2	F90495
111	7	0.9	305	2	F84595
112	7	0.9	318	2	G87342
113	7	0.9	319	2	G75106
114	7	0.9	322	2	H83851
115	7	0.9	323	2	A54021
116	7	0.9	323	2	B54021
117	7	0.9	327	1	OKDRC
118	7	0.9	327	2	A55545
119	7	0.9	331	2	T48867
120	7	0.9	333	1	J50590
121	7	0.9	338	2	S26976
122	7	0.9	338	2	J10121
123	7	0.9	340	2	T35394
124	7	0.9	343	2	G87995
125	7	0.9	344	2	A97699
126	7	0.9	344	2	AB2925
127	7	0.9	347	2	G87449
128	7	0.9	372	2	AG1896
129	7	0.9	373	2	F84396
130	7	0.9	379	2	F71533
131	7	0.9	381	2	H69367
132	7	0.9	383	2	S51327
133	7	0.9	384	2	D69411
134	7	0.9	386	1	KMECPW
135	7	0.9	386	2	B85906
136	7	0.9	386	2	F91061
137	7	0.9	387	2	H65132
138	7	0.9	399	2	C87387
139	7	0.9	399	2	F87547
140	7	0.9	402	2	H86649
141	7	0.9	405	2	H88930
142	7	0.9	405	2	A46699
143	7	0.9	406	2	H70670
144	7	0.9	415	2	S63221
145	7	0.9	416	2	D72456
146	7	0.9	423	2	D84021
147	7	0.9	423	2	S781248
148	7	0.9	424	2	T15959
149	7	0.9	425	2	G72456
150	7	0.9	430	2	C84190
151	7	0.9	432	2	B83891
152	7	0.9	444	2	T31581
153	7	0.9	445	2	UBZPG
154	7	0.9	446	1	G84001
155	7	0.9	447	2	S00950
156	7	0.9	450	2	UBHUG
157	7	0.9	451	1	UBXUG
158	7	0.9	451	1	UBXUG
159	7	0.9	452	2	A10779
160	7	0.9	453	2	A81665
161	7	0.9	455	2	T15622
162	7	0.9	455	2	B96037
163	7	0.9	457	2	J01413
164	7	0.9	457	2	A35356
165	7	0.9	461	1	S53084
166	7	0.9	462	2	G87534
167	7	0.9	462	2	B91186
168	7	0.9	466	2	A86033
169	7	0.9	466	2	S01943
170	7	0.9	467	2	JC5798
171	7	0.9	469	2	S01937
172	7	0.9	471	2	VGBEF2
173	7	0.9	479	1	A56694
174	7	0.9	480	2	
175	7	0.9			

tropomyosin - b100
 tropomyosin - Cali
 probable lipase ac
 hypothetical prote
 mannose/glucose-bi
 Lyat family transc
 conserved hypotet
 conserved hypotet
 hypothetical prote
 hypothetical prote
 hypothetical prote
 hypothetical prote
 phosphotyrosyl pho
 phosphotyrosyl pho
 F1pi protein - yea
 dtdglucose 4,6-de
 endo-1,4-beta-xy1a
 glyceraldehyde-3-p
 glyceraldehyde-3-p
 probable glycerol
 hypothetical prote
 utrophyrinogen d
 conserved hypotet
 hypothetical prote
 signal recognition
 hypothetical prote
 hypothetical prote
 heparin-binding gl
 probable acyl-CoA
 chorismate mutase
 chorismate mutase-
 chorismate mutase-
 hypothetical 44.3
 beta-ketoadipyl-Co
 beta-ketoadipyl-Co
 rhamnosyltransfera
 protein R1G11.14
 translation initia
 hypothetical prote
 DNA-directed RNA p
 probable glutamyl-
 3-deoxy-D-manno-oc
 3-deoxy-D-manno-oc
 hypothetical prote
 hypothetical prote
 cytophan synthas
 probable glutamate
 intracellular alka
 hypothetical prote
 tubulin gamma chai
 glycine dehydrogen
 hypothetical prote
 tubulin gamma chai
 tubulin gamma chai
 probable n-hydroxy
 hypothetical RNA m
 hypothetical prote
 oxidoreductase (lm
 probable alpha-gal
 hypothetical 50.9K
 tumor necrosis fac
 gamma-tubulin - Bu
 UDP-N-acetylglucos
 hypothetical prote
 probable permeal
 probable permeal
 F0F1-ATPase (EC 3.
 hypothetical prote
 glycoprotein F - h
 keratin, type II,

176	7	0.9	483	2	T36811	probable integral
177	7	0.9	488	1	S55874	sulfite oxidase (E
178	7	0.9	490	2	B86813	xyloside transport
179	7	0.9	491	2	S05408	keratin, type II,
180	7	0.9	493	2	H83052	N utilization subs
181	7	0.9	499	2	G75600	cleavage and poly
182	7	0.9	503	2	T24985	hypothetical prote
183	7	0.9	503	2	S29094	keratin, type II,
184	7	0.9	508	1	KRSHL2	keratin type II, m
185	7	0.9	508	2	S41877	readthrough protei
186	7	0.9	508	2	S41879	readthrough protei
187	7	0.9	508	2	S41875	readthrough protei
188	7	0.9	508	2	S03551	hypothetical prote
189	7	0.9	508	2	S24594	hypothetical prote
190	7	0.9	510	2	C97232	GMP synthase (limp
191	7	0.9	512	1	TVHUX	GMP synthase (limp
192	7	0.9	512	2	G91200	protein-tyrosine k
193	7	0.9	512	2	G91200	type III secretion
194	7	0.9	512	2	180311	sepc protein - Esc
195	7	0.9	520	2	C86047	escC (imported) -
196	7	0.9	520	2	T04591	ferulate-5-hydroxy
197	7	0.9	524	2	C37222	cytochrome P450 1A
198	7	0.9	534	2	S31300	regulatory protein
199	7	0.9	538	2	S67766	RNA-export mediat
200	7	0.9	540	2	AG2161	GMP synthase (Gut
	7	0.9	542	2	S76358	GMP synthase (Gut

ALIGNMENTS

RESULT 1

C95136 conserved domain protein SP1175 (imported) - Streptococcus pneumoniae (strain TIGR4)

C:Species: Streptococcus pneumoniae

C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001

C:Accession: C95136

C:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

C:Accession: C95136

C:Status: preliminary

A:Residues: 1-802 <KOR>

A:Cross-references: GB:AE005672; PIDN:AAK75284.1; PID:G14972655; GSPDB:GN00164; TIGR:SP4

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SP1175

C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 87.3%; Score 695; DB 2; Length 802;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	SYELGLVQARTVKENNRVSVDGKQATOKTENLTPDEVSREGINAOIVIKIDOGVYT	60
DB	7	SYELGLVQARTVKENNRVSVDGKQATOKTENLTPDEVSREGINAOIVIKIDOGVYT	66
QY	61	SHGHHYNGKGVYDAIISFELMKDPNYKLKDEIVNEVKGGVYIKVDKGVYVYLKDA	120
DB	67	SHGHHYNGKGVYDAIISFELMKDPNYKLKDEIVNEVKGGVYIKVDKGVYVYLKDA	126
QY	121	AHADNVATKEEINRQKHSGHREGGTPRNDGAVALARSQGRVYTTDDGYIFNA3DIIDT	180
DB	127	AHADNVATKEEINRQKHSGHREGGTPRNDGAVALARSQGRVYTTDDGYIFNA3DIIDT	186
QY	181	GDAIVVPHGHYHYIIPKNELSASFLAAEAFLSGRGLNSRTYRRQNSDTSRTNWPS	240
DB	187	GDAIVVPHGHYHYIIPKNELSASFLAAEAFLSGRGLNSRTYRRQNSDTSRTNWPS	246

QY 241 VSNGETNTNTSNTNSQASQASNDIDSLKQLYKPLSQRHVESDGLVFDPQAQITSTR 300
DB 247 VSNPGTTNTNTSNTNSQASQASNDIDSLKQLYKPLSQRHVESDGLVFDPQAQITSTR 306
QY 301 AGVAAPHGDHFIYISQMSLEERLARIIPLRYSNMHWPPDSRPEQSPQPTPEPSRQ 360
DB 307 AKVAAPHGDHFIYISQMSLEERLARIIPLRYSNMHWPPDSRPEQSPQPTPEPSRQ 366
QY 361 POPAPNLKIDSNSLSQVLRKYGEGVFEKGISRVFAKDPSETVKLEKLSKQES 420
DB 367 POPAPNLKIDSNSLSQVLRKYGEGVFEKGISRVFAKDPSETVKLEKLSKQES 426
QY 421 VSHITLAKKENVAPRDOEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLERLNDESTN 480
DB 427 VSHITLAKKENVAPRDOEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLERLNDESTN 486
QY 481 KEKLVDDLAFAPITHPERLGRPNQIETDEVRIAQLADKYTTSDDGIYFDEHDIISD 540
DB 487 KEKLVDDLAFAPITHPERLGRPNQIETDEVRIAQLADKYTTSDDGIYFDEHDIISD 546
QY 541 EGDAYVTPHMGSHWIGKDSLSDKEKYAAQAYTKERGIIPSPDADVKANPTGDSAAAY 600
DB 547 EGDAYVTPHMGSHWIGKDSLSDKEKYAAQAYTKERGIIPSPDADVKANPTGDSAAAY 606
QY 601 NRVKGEKRIPLVRLPYMVEHTVEVKNGNLIIPKDHYNIKFAMPDHTYKAPNGYTLBD 660
DB 607 NRVKGEKRIPLVRLPYMVEHTVEVKNGNLIIPKDHYNIKFAMPDHTYKAPNGYTLBD 666
QY 661 LPATIKYVEHPERHSHNDGCMASBHVGLKQHSBEPKPKFADBEPEVETPAEPV 720
DB 667 LPATIKYVEHPERHSHNDGCMASBHVGLKQHSBEPKPKFADBEPEVETPAEPV 726
QY 721 QVETEKYEAQLKEAEVLAKVTDSSLKANATETLAGLRNNLTLOIMDNNSIMAEKELIA 780
DB 727 QVETEKYEAQLKEAEVLAKVTDSSLKANATETLAGLRNNLTLOIMDNNSIMAEKELIA 786
QY 781 LKGSNPSVSKEKIN 796
DB 787 LKGSNPSVSKEKIN 802

RESULT 2

E98004
hypothetical protein pnta [imported] - Streptococcus pneumoniae (strain R6)
C/Species: Streptococcus pneumoniae
C/Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #ext_change 02-Nov-2001
C/Accession: E98004
R./Hoskins, J.A.; Albom Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.; E
e, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
Y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A./Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.
A./Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A./Reference number: A97872; PMID:21429245; PMID:11544234
A/Accession: E98004
A/Status: preliminary
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-828 <KUR>
A/Cross-references: GB:AE007317; PIDN:AAK99865.1; PID:G15458683; GSPDB:GN00174
A/Genes: pnta
C/Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 77.4%; Score 616; DB 2; Length 828;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 716; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 80 SEELAKDPYKLDIDIVNEVKGYYIKVDGYYVYLKDAHADNVRTKEEINRQKOE 139
DB 112 SEELAKDPYKLDIDIVNEVKGYYIKVDGYYVYLKDAHADNVRTKEEINRQKOE 171
QY 140 SQREGGTPRNDGAVALARSQGRYTTDDGYIFNASDIIEDTDGAYIVPHGDHYHYI PKNE 199

DB 172 SQREGGTPRNDGAVALARSQGRYTTDDGYIFNASDIIEDTDGAYIVPHGDHYHYI PKNE 231
QY 200 LSASELAEEAFPLSGRGNLSNGRTYRRONSDNTSRNTWPSVNGCTNTNTSNTNS 259
DB 232 LSASELAEEAFPLSGRGNLSNGRTYRRONSDNTSRNTWPSVNGCTNTNTSNTNS 291
QY 260 QASQNDIDSLKQLYKPLSQRHVESDGLVFDPQAQITSTRAGVAPHGDHFIYISQ 319
DB 292 QASQNDIDSLKQLYKPLSQRHVESDGLVFDPQAQITSTRAGVAPHGDHFIYISQ 351
QY 320 MSELERLARIIPLRYSNMHWPPDSRPEQSPQPTPEPSRQPOAPNLKIDSNSLSQ 379
DB 352 MSELERLARIIPLRYSNMHWPPDSRPEQSPQPTPEPSRQPOAPNLKIDSNSLSQ 411
QY 380 VRVKGEYVFEKGISRVFAKDPSETVKNLKESKQESVSHITLAKKENVAPRDOEF 439
DB 412 VRVKGEYVFEKGISRVFAKDPSETVKNLKESKQESVSHITLAKKENVAPRDOEF 471
QY 440 YRKAVNLLTEAHKALFXNKGNSDFQALDKLERLNDESTNKEKLVDDLAFAPITHP 499
DB 472 YRKAVNLLTEAHKALFXNKGNSDFQALDKLERLNDESTNKEKLVDDLAFAPITHP 531
QY 500 RLKGNPQIETDEVRIAQLADKYTTSDDGIYFDEHDIISDEGDAYVTPHMGSHWIGK 559
DB 532 RLKGNPQIETDEVRIAQLADKYTTSDDGIYFDEHDIISDEGDAYVTPHMGSHWIGK 591
QY 560 SLSDKEKYAAQAYTKERGIIPSPDADVKANPTGDSAAAYNRVGEKRIPLVRLPYME 619
DB 592 SLSDKEKYAAQAYTKERGIIPSPDADVKANPTGDSAAAYNRVGEKRIPLVRLPYME 651
QY 620 HTVEVANGNLIIPKDHYNIKFAMPDHTYKAPNGYTLBDLPATIKYVEHPERHSHN 679
DB 652 HTVEVANGNLIIPKDHYNIKFAMPDHTYKAPNGYTLBDLPATIKYVEHPERHSHN 711
QY 680 DGMGNASEHVGLKQHSBEPKPKFADBEPEVETPAEPVQVETEKYEAQLKEAEVLIA 739
DB 712 DGMGNASEHVGLKQHSBEPKPKFADBEPEVETPAEPVQVETEKYEAQLKEAEVLIA 771
QY 740 KYTDSILKANATETLAGLRNNLTLOIMDNNSIMAEKELIALKGSNPSVSKEKIN 796
DB 772 KYTDSILKANATETLAGLRNNLTLOIMDNNSIMAEKELIALKGSNPSVSKEKIN 828

RESULT 3

D98004
histidine Motif-Containing protein [imported] - Streptococcus pneumoniae (strain R6)
C/Species: Streptococcus pneumoniae
C/Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #ext_change 02-Nov-2001
C/Accession: D98004
R./Hoskins, J.A.; Albom Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.; E
e, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
Y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A./Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.
A./Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A./Reference number: A97872; PMID:21429245; PMID:11544234
A/Accession: D98004
A/Status: preliminary
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-855 <KUR>
A/Cross-references: GB:AE007317; PIDN:AAK99864.1; PID:G15458682; GSPDB:GN00174
A/Genes: pnta
C/Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 13.4%; Score 107; DB 2; Length 855;
Best Local Similarity 100.0%; Pred. No. 1.8e-99;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 102 KGGYIKVDGKYVYLKDAHADNVRTKEEINRQKOEHSQREGGTPRNDGAVALARSQ 161
DB 133 KGGYIKVDGKYVYLKDAHADNVRTKEEINRQKOEHSQREGGTPRNDGAVALARSQ 192

A:Molecule type: DNA
A:Residues: 1-1039 <KUR>
A:Cross-References: GB:AE007317; PIDN:AAK99712.1; PID:g15458515; GSPDB:GN00174
C:Genetics:
A:Gene: phcE

Query Match 4.0%; Score 32; DB 2; Length 1039;
Best Local Similarity 100.0%; Pred. No. 2.3e-23;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 AEOIVKIDGCVTSHGDHYHYNGKVPYDA 77
DB 67 AEOIVKIDGCVTSHGDHYHYNGKVPYDA 98

RESULT 9

T46758
hypothetical 92.4K protein - Streptococcus agalactiae

C:Species: Streptococcus agalactiae

C:Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 21-Jul-2000

C:Accession: T46758

R:Spellerberg, B.; Rozdzinski, E.; Martin, S.; Weber-Heynemann, J.; Schuttzler, N.; Luet

Infect. Immun. 67, 871-878, 1999

A:Title: Lmb, a protein with similarities to the Lrai adhesin family, mediates attachment

A:Reference number: Z24091; MUID:99115568; PMID:9916102

A:Accession: T46758

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-822 <SPB>

A:Cross-References: EMBL:AF062533; NID:g4249622; PIDN:AAD13797.1; PID:g4249624

A:Experimental source: strain R268

C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 2.1%; Score 17; DB 2; Length 822;
Best Local Similarity 100.0%; Pred. No. 3e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YNGKVPYDAIISEELIM 85
DB 92 YNGKVPYDAIISEELIM 108

RESULT 10

F97985

hypothetical protein phcE-truncation [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae

C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001

C:Accession: F97985

R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.; F

e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M

J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;

A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A:Reference number: A97872; MUID:21429245; PMID:11544234

A:Accession: F97985

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-182 <KUR>

A:Cross-References: GB:AE007317; PIDN:AAK99714.1; PID:g15458517; GSPDB:GN00174

C:Genetics:
A:Gene: phcE-truncation

Query Match 1.3%; Score 10; DB 2; Length 182;
Best Local Similarity 100.0%; Pred. No. 0.096;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 188 HGDHYHYIPK 197
DB 35 HGDHYHYIPK 44

RESULT 11

A48184
transcription initiation factor IID 110K chain - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C:Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 15-Oct-1999

C:Accession: A48184

R:Kokubo, T.; Gong, D.W.; Roeder, R.G.; Horikoshi, M.; Nakatani, Y.

Proc. Natl. Acad. Sci. U.S.A. 90, 5896-5900, 1993

A:Title: The Drosophila 110-kDa transcription factor TFIID subunit directly interacts wi

A:Reference number: A48184; MUID:93117591; PMID:8327460

A:Accession: A48184

A:Status: preliminary

A:Molecule type: mRNA; protein

A:Residues: 1-921 <KOK>

A:Cross-References: GB:S63550; NID:g398432; PID:g398433

A:Experimental source: embryo nuclear extract

A:Note: sequence extracted from NCBI backbone (NCBIN:134863; NCBI:134864)

C:Genetics:
A:Gene: FlyBase:Tafl10

A:Cross-References: FlyBase:FBgn0010280

C:Keywords: transcription initiation

Query Match 1.1%; Score 9; DB 2; Length 921;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 TTNNSNSNT 257
DB 178 TTNNSNSNT 186

RESULT 12

A45183

TBP-associated factor TFIID - fruit fly (Drosophila sp.)

C:Species: Drosophila sp.

C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 16-Feb-1997

C:Accession: A45183

R:Hoey, T.; Weinzierl, R.O.; Gill, G.; Chen, J.L.; Dynlacht, B.D.; Tjian, R.

Cell 72, 247-260, 1993

A:Title: Molecular cloning and functional analysis of Drosophila TAF110 reveal propertie

A:Reference number: A45183; MUID:93145326; PMID:7678780

A:Accession: A45183

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-921 <HOE>

A:Experimental source: embryo

A:Note: sequence extracted from NCBI backbone (NCBIP:123832)

C:Genetics:
A:Gene: FlyBase:Tafl10

A:Cross-References: FlyBase:FBgn0010280

Query Match 1.1%; Score 9; DB 2; Length 921;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 TTNNSNSNT 257
DB 178 TTNNSNSNT 186

RESULT 13

E72853

AcOrf-29 protein - Autographa californica nuclear polyhedrosis virus

C:Species: Autographa californica nuclear polyhedrosis virus, AcMNPV

A:Note: dsDNA virus

C:Date: 12-Nov-1999 #sequence_revision 12-Nov-1999 #text_change 12-Nov-1999

C:Accession: E72853

R:AYres, M.D.; Howard, S.C.; Kuzio, J.; Lopez-Ferber, M.; Possee, R.D.

Virology 202, 586-605, 1994

A:Title: The complete DNA sequence of Autographa californica nuclear polyhedrosis virus

A:Reference number: A72850; MUID:94303173; PMID:8030224

A:Accession: E72853

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-71 <AYR>
A:Cross-references: GB:L22858; NID:G510708; PIDN:AAA6659.1; PID:G559098
C:Genetics:
A:Gene: ACO-f-29

Query Match
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 130 EFINRQKQ 137
DB 20 EFINRQKQ 27

RESULT 14
S58921
tropomyosin isoform TMI - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
C:Date: 15-Feb-1996 #sequence revision 01-Mar-1996 #text_change 20-Jun-2000
C:Accession: S58921; S62127; S58918
R:Kagawa, H.; Sugimoto, K.; Matsumoto, H.; Inoue, T.; Imadzu, H.; Takuwa, K.; Sakube, Y.
J. Mol. Biol. 251, 603-613, 1995
A>Title: Genome structure, mapping and expression of the tropomyosin gene tmy-1 of Caenc
A:Reference number: S58918; MUID:95395840; PMID:7666414
A:Accession: S58921

A:Molecule type: mRNA
A:Residues: 1-284 <KAG>
A:Cross-references: EMBL:D38540; NID:G1208412; PIDN:BA07543.1; PID:G1208413
A:Accession: S62127
A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
A:Residues: 1-276, 'NLP', 280-284 <KAF>
A:Cross-references: EMBL:D38539; NID:G871835; PIDN:BA07540.1; PID:G1208409
A:Accession: S58918
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1994
A:Molecule type: protein
A:Residues: 11-49; 171-211 <KAW>
C:Genetics:

A:Gene: tmy-1
A:introns: 15/3; 61/3; 80/3; 125/2; 164/3; 188/2; 234/3; 258/1
C:Superfamily: tropomyosin
C:Keywords: alternative splicing

Query Match
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 727 VEAQLKEA 734
DB 144 VEAQLKEA 151

RESULT 15
S58922

tropomyosin isoform TMI - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
C:Date: 10-Apr-1996 #sequence revision 19-Apr-1996 #text_change 20-Jun-2000
C:Accession: S58922; S62128; S58918
R:Kagawa, H.; Sugimoto, K.; Matsumoto, H.; Inoue, T.; Imadzu, H.; Takuwa, K.; Sakube, Y.
J. Mol. Biol. 251, 603-613, 1995
A>Title: Genome structure, mapping and expression of the tropomyosin gene tmy-1 of Caenc
A:Reference number: S58918; MUID:95395840; PMID:7666414
A:Accession: S58922

A:Molecule type: mRNA
A:Residues: 1-284 <KAG>
A:Cross-references: EMBL:D38541; NID:G871837; PIDN:BA07544.1; PID:G1208414
A:Accession: S62128
A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
A:Residues: 1-284 <KAW>
A:Cross-references: EMBL:D38539; NID:G871835; PIDN:BA07541.1; PID:G1208410
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1994
A:Accession: S58918

A:Molecule type: protein
A:Residues: 11-49; 171-211 <KAF>
C:Genetics:

A:Gene: tmy-1
A:introns: 15/3; 61/3; 80/3; 125/2; 164/3; 188/2; 234/3; 258/1
C:Superfamily: tropomyosin
C:Keywords: alternative splicing

Query Match
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 727 VEAQLKEA 734
DB 144 VEAQLKEA 151

RESULT 16
T49361

probable transport vesicle fusion protein SEC17 [imported] - Neurospora crassa

N:Alternate names: protein B1D.150
C:Species: Neurospora crassa
C:Date: 02-Jun-2000 #sequence revision 02-Jun-2000 #text_change 02-Jun-2000
C:Accession: T49361
R:Schulte, U.; Align, V.; Hohnsbeil, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, May 2000
A:Reference number: 225022
A:Accession: T49361

A>Status: preliminary
A:Molecule type: DNA

A:Residues: 1-292 <SCH>
A:Cross-references: EMBL:AL355927; GSPDB:GN00116; NCSP:B1D1.150

A:Experimental source: BAC clone B1D1, strain OR74A
C:Genetics:

A:Gene: NCSP:B1D1.150
A:Map position: 6

A:introns: 11/3; 53/1; 155/2

Query Match
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 NDGAVALA 157
DB 150 NDGAVALA 157

RESULT 17
T27791

hypothetical protein ZK228.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence revision 15-Oct-1999 #text_change 20-Jun-2000
C:Accession: T27791
R:Baaham, V.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z20419
A:Accession: T27791

A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

A:Residues: 1-294 <WLL>
A:Cross-references: EMBL:Z82086; PIDN:CAB04996.1; GSPDB:GN00023; CESP:ZK228.4

A:Experimental source: clone ZK228
C:Genetics:

A:Gene: CESP:ZK228.4
A:Map position: 5

A:introns: 38/1; 99/1; 143/3; 181/3; 250/2
C:Superfamily: Caenorhabditis elegans hypothetical protein F31F7.1

Query Match
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 TVKENNRV 18

DB 262 TVKENNRV 269
|||||
RESULT 18
S05355
hypothetical protein (clone AAC11) - slime mold (Dictyostellium discoideum) (fragment)
C/Species: Dictyostellium discoideum
C/Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 29-Oct-1999
C/Accession: S05355
R/Shaw, D.R.; Richter, H.; Giorda, R.; Ohmachi, T.; Ennis, H.L.
Mol. Gen. Genet. 218, 453-459, 1989
A/Title: Nucleotide sequences of Dictyostellium discoideum developmentally regulated cdna
A/Reference number: S05355; MUID:90066348; PMID:2511421
A/Accession: S05355
A/Molecule type: mRNA
A/Residues: 1-448 <SHA>
A/Cross-references: EMBL:X16522; NID:g7173; PIDD:CAA34529.1; PID:g7174
Query Match
1.0%; Score 8; DB 2; Length 448;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 250 NTSNNSNT 257
|||||
DB 378 NTSNNSNT 385
RESULT 19
F82203
probable multidrug resistance protein VC1409 [imported] - Vibrio cholerae (strain N16961)
C/Species: Vibrio cholerae
C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C/Accession: F82203
R/Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Glimm, M.L.; Dodson, R.U.;
Chardon, D.; Ermolaeva, M.D.; Vamathevan, J.; Baas, S.; Qin, H.; Dragol, I.; Sellers, F.
L.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A/Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A/Reference number: AB2035; MUID:20406833; PMID:10952301
A/Accession: F82203
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-484 <HEI>
A/Cross-references: GB:AE004219; GB:AE003852; NID:G9655899; PIDD:AAF94566.1; GSPDB:GN001
C/Genetics:
A/Experimental source: serogroup O1, strain N16961, biotype El Tor
A/Map position: 1
Query Match
1.0%; Score 8; DB 2; Length 484;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 203 SELAAEA 210
|||||
DB 169 SELAAEA 176
RESULT 20
B89957
hypothetical protein ptaA [imported] - Staphylococcus aureus (strain N315)
C/Species: Staphylococcus aureus
C/Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C/Accession: B89957
R/Ruoda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguro
ma, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A/Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A/Reference number: AB9758; MUID:21311952; PMID:11418146
A/Accession: B89957
A/Status: preliminary

A/Molecule type: DNA
A/Residues: 1-488 <KUR>
A/Cross-references: GB:BA000018; PID:g13701521; PIDD:BA42815.1; GSPDB:GN00149
A/Experimental source: strain N315
C/Genetics:
A/Gene: ptaA
C/Superfamily: phosphotransferase system glucose-specific enzyme II, factor II; phosphot
Query Match
1.0%; Score 8; DB 2; Length 488;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 150 NDGAVALA 157
|||||
DB 72 NDGAVALA 79
RESULT 21
S73830
endopeptidase Ia (EC 3.4.21.53) - Mycoplasma pneumoniae (strain ATCC 29342)
N/Alternate names: ATP-dependent proteinase Ipn; ATP-dependent serine proteinase Ia; hyl
N:contains: adenosinetriphosphatase (EC 3.6.1.3)
C/Species: Mycoplasma pneumoniae
A/Variety: ATCC 29342
C/Date: 09-Apr-1999 #sequence_revision 09-Apr-1999 #text_change 19-Jan-2001
C/Accession: S73830
R/Himmelsreich, R.; Hilbert, H.; Plagens, H.; Pirk, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A/Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae
A/Reference number: S73327; MUID:97105885; PMID:8948633
A/Accession: S73830
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-795 <HIM>
A/Cross-references: EMBL:AE000050; GB:U00089; NID:g1674197; PIDD:AAB96152.1; PID:g167419
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
C/Comment: This enzyme catalyzes the hydrolysis of large proteins in the presence of ATP
C/Genetics:
A/Gene: Ipn; F10 orf795
A/Genetic code: EGC3
C/Superfamily: ATP-dependent serine proteinase Ia
C/Keywords: ATP; DNA binding; heat shock; hydrolase; nucleotide binding; P-loop; serine
F;379-386/Region: nucleotide-binding motif A (P-loop)
F;442-447/Region: nucleotide-binding motif B
F;702/Active site: Ser #status predicted
Query Match
1.0%; Score 8; DB 1; Length 795;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 466 ALDKLLER 473
|||||
DB 127 ALDKLLER 134
RESULT 22
T00818
hypothetical protein At2g41620 [imported] - Arabidopsis thaliana
N/Alternate names: hypothetical protein T32G6.14
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 23-Mar-2001
C/Accession: T00818; AB4844
R/Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
submitted to the EMBL Data Library, November 1997
A/Description: Arabidopsis thaliana chromosome II BAC T32G6 genomic sequence.
A/Reference number: Z14163
A/Accession: T00818
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-825 <ROU>
A/Cross-references: EMBL:AC002510; NID:g2618683; PID:g2618698
A/Experimental source: cultivar Columbia
R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koc, H.; Mofatt, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.; Neuss, D.; Niemann, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999

A>Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.

A.Reference number: AB4420; MUID:20083487; PMID:1061197

A.Accession: AB4844

A.Status: preliminary

A.Molecule type: DNA

A.Residues: 1-825 <STO>

A.Cross-references: GB:AE002093; NID:g2618698; PIDN:AA84345.1; GSPDB:GN00139

C.Genetics:

A.Gene: At2g41620; T32G6.14

A.Map position: 2

A.Introns: 43/3; 74/2; 96/3; 116/3; 146/3; 234/3; 311/3; 329/3; 368/3; 669/3; 725

A.Note: T32G6.14

C.Superfamily: Arabidopsis thaliana hypothetical protein At2g41620

Query Match

Best Local Similarity 1.0%; Score 8; DB 2; Length 825;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 REGINAEO 48

Db 78 REGINAEO 85

RESULT 23

C75313

probable proteinase - *Deinococcus radiodurans* (strain R1)

C.Species: *Deinococcus radiodurans*

C.Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000

C.Accession: C75313

R.White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; S. Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A>Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.

A.Reference number: A75250; MUID:20036896; PMID:10567266

A.Accession: C75313

A.Status: preliminary

A.Molecule type: DNA

A.Residues: 1-840 <WHI>

A.Cross-references: GB:AE002047; GB:AE000513; NID:g6459915; PIDN:AA11680.1; PID:g645992

A.Experimental source: strain R1

C.Genetics:

A.Gene: DR2130

A.Map position: 1

Query Match

Best Local Similarity 1.0%; Score 8; DB 2; Length 840;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 287 DGLVDPDA 294

Db 332 DGLVDPDA 339

RESULT 24

A64714

helicase - *Helicobacter pylori* (strain 26695)

C.Species: *Helicobacter pylori*

C.Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999

C.Accession: A64714

R.Tomb, J.F.; White, O.; Kertavagie, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L. Nature 388, 533-547, 1997

A.Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpe, P.D.; Smith, H.O.; Fraser, C.

A>Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.

A.Reference number: A64520; MUID:97394467; PMID:9252185

A.Accession: A64714

A.Status: preliminary; nucleic acid sequence not shown; translation not shown

A.Molecule type: DNA

A.Residues: 1-945 <TOM>

A.Cross-references: GB:AE000653; GB:AE000511; NID:g2314733; PIDN:AD08593.1; PID:g231473

Query Match

Best Local Similarity 1.0%; Score 8; DB 2; Length 945;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 499 ERLGKPN 506

Db 657 ERLGKPN 664

RESULT 25

B70748

probable smc protein - *Mycobacterium tuberculosis* (strain H37RV)

C.Species: *Mycobacterium tuberculosis*

C.Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C.Accession: B70748

R.Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamilton, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998

A.Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A>Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome

A.Reference number: A70500; MUID:98295987; PMID:9634230

A.Accession: B70748

A.Status: preliminary; nucleic acid sequence not shown; translation not shown

A.Molecule type: DNA

A.Residues: 1-1289 <COL>

A.Cross-references: GB:B74697; GB:AL123456; NID:g3361602; PIDN:CAA98982.1; PID:g1405958

A.Experimental source: strain H37RV

C.Genetics:

A.Gene: smc

C.Superfamily: chromosome segregation protein SMCI

Query Match

Best Local Similarity 1.0%; Score 8; DB 2; Length 1289;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 203 SELAAEA 210

Db 756 SELAAEA 763

RESULT 26

H97926

hypothetical protein spv0440 [imported] - *Streptococcus pneumoniae* (strain R6)

C.Species: *Streptococcus pneumoniae*

C.Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001

C.Accession: H97926

R.Hocking, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E. R.; LeBlanc, D.J.; Lee, E.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAnren, S.; M. Y. P.; Sun, P.M.; Winkler, M.E. J. Bacteriol. 183, 5709-5717, 2001

A.Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.

A>Title: Genome of the Bacterium *Streptococcus pneumoniae* Strain R6.

A.Reference number: A97872; MUID:21429245; PMID:11544224

A.Accession: H97926

A.Status: preliminary

A.Molecule type: DNA

A.Residues: 1-1659 <KUR>

A.Cross-references: GB:AE007317; PIDN:AAK9244.1; PID:g15458007; GSPDB:GN00174

C.Genetics:

A.Gene: spv0440

Query Match

Best Local Similarity 1.0%; Score 8; DB 2; Length 1659;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 447 LTEAHKAL 454

Db 1292 LTEAHKAL 1299

RESULT 27
G95057
endo-beta-N-acetylglucosaminidase, probable [imported] - Streptococcus pneumoniae (strain
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #ext_change 03-Aug-2001
C:Accession: G95057
R:Rettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,
mon, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.
A:Reference number: A95000; PMID:21357209; PMID:11463916
A:Accession: G95057
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1659 <KUR>
A:Cross-references: GB:A0005672; PIDN:AAK74656.1; PID:g14971970; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP0498
Query Match 1.0%; Score 8; DB 2; Length 1659;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 447 LFEAKAL 454
| | | | | | | |
Db 1292 LFEAKAL 1299
RESULT 28
SS5073
env polyprotein - simian immunodeficiency virus STVmac (strain macC8) (fragment)
C:Species: simian immunodeficiency virus STVmac
A:Variety: strain macC8
C:Date: 08-Jul-1995 #sequence_revision 21-Jul-1995 #ext_change 26-Aug-1999
C:Accession: SS5073
R:Whitmore, A.; Cook, N.; Hall, G.; Sharpe, S.; Rud, E.; Cranage, M.
submitted to the EMBL Data Library, April 1995
A:Description: Repair and evolution of nef in vivo modulates SIV virulence.
A:Reference number: S54796
A:Accession: SS5073
A:Molecule type: DNA
A:Residues: 1-44 <WHA>
A:Cross-references: EMBL:X86728; NID:g840891; PIDN:CAA60402.1; PID:g861124
A:Experimental source: strain macC8; week 33 after infection
C:Genetics:
A:Gene: env
C:Superfamily: type E retrovirus env polyprotein
C:Keywords: AIDS; immunodeficiency; polyprotein
Query Match 0.9%; Score 7; DB 2; Length 44;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 750 ATEETLAG 756
| | | | | | | |
Db 6 ATEETLAG 12
RESULT 29
SS5072
env polyprotein - simian immunodeficiency virus STVmac (strain macC8) (fragment)
C:Species: simian immunodeficiency virus STVmac
A:Variety: strain macC8
C:Date: 08-Jul-1995 #sequence_revision 21-Jul-1995 #ext_change 26-Aug-1999
C:Accession: SS5072
R:Whitmore, A.; Cook, N.; Hall, G.; Sharpe, S.; Rud, E.; Cranage, M.
submitted to the EMBL Data Library, April 1995
A:Description: Repair and evolution of nef in vivo modulates SIV virulence.
A:Reference number: S54796

A:Accession: SS5072
A:Molecule type: DNA
A:Residues: 1-49 <WHA>
A:Cross-references: EMBL:X86727; NID:g840890; PIDN:CAA60400.1; PID:g861122
A:Experimental source: strain macC8; week 25 after infection
C:Genetics:
A:Gene: env
C:Superfamily: type E retrovirus env polyprotein
C:Keywords: AIDS; immunodeficiency; polyprotein
Query Match 0.9%; Score 7; DB 2; Length 49;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 750 ATEETLAG 756
| | | | | | | |
Db 11 ATEETLAG 17
RESULT 30
D70091
hypothetical protein ynfF - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #ext_change 15-Oct-1999
C:Accession: D70091
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Beretti
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Erttington, J.; Fabret, C.; Ferrati, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
Y, M.; Ogasawa, K.; Ogawa, A.; Oudega, B.; Park, S.H.; Paro, V.; Pohl, T.M.; Portet
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A:Authors: Schleich, S.; Schreier, R.; Scofield, F.; Sekiguchi, J.; Sekowska, A.; Sero
akuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wilpat, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumberg, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; PMID:98044033; PMID:9384377
A:Accession: D70091
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-49 <KUN>
A:Cross-references: GB:Z99124; GB:A0009126; NID:g2636442; PIDN:CAB16055.1; PID:el184744,
A:Experimental source: strain 168
C:Genetics:
A:Gene: ynfF
Query Match 0.9%; Score 7; DB 2; Length 49;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 406 ETVKNLE 412
| | | | | | | |
Db 9 ETVKNLE 15
RESULT 31
D87657
hypothetical protein CC3294 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #ext_change 20-Apr-2001
C:Accession: D87657
R:Nierman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; PMID:21173698; PMID:11259647
A:Accession: D87657
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-68 <STO>
A:Cross-references: GB:AE005673; NID:g13424988; PIDN:AAK5256.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC3294

Query Match 0.9%; Score 7; DB 2; Length 68;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 322 EMBEIRA 328
|||||
Db 31 EMBEIRA 37

RESULT 32

A75134
Hypothetical protein PAB8218 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: A75134
R:anonymous, Genoscope
Submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
A:Reference number: A75001
A:Accession: A75134
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-70 <RAW>
A:Cross-references: GB:AJ248285; GB:AL096836; NID:gs458067; PIDN:CAB49786.1; PID:el51568
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB8218

Query Match 0.9%; Score 7; DB 2; Length 70;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 478 STNKEKL 484
|||||
Db 34 STNKEKL 40

RESULT 33

C90404
Conserved hypothetical protein [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C:Accession: C90404
R:She, O.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, F.
Submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: C90404
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-90 <RUR>
A:Cross-references: GB:AE006641; NID:g13815638; PIDN:AAK42490.1; GSPDB:GN00155
C:Genetics:
A:Gene: SSO10237

Query Match 0.9%; Score 7; DB 2; Length 90;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 GKYYVYL 117
|||||
Db 18 GKYYVYL 24

RESULT 34

C75396
Hypothetical protein - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: C75396
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, H.O.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma-
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266

A:Accession: C75396
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-93 <WHI>
A:Cross-references: GB:AE001987; GB:AE00513; NID:g6459180; PIDN:AAF11000.1; PID:g645918
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1421
A:Map position: 1

Query Match 0.9%; Score 7; DB 2; Length 93;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 487 DLAFILA 493
|||||
Db 6 DLAFILA 12

RESULT 35

E75556
Hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: E75556
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, H.O.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma-
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: E75556
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-94 <WHI>
A:Cross-references: GB:AE001875; GB:AE00513; NID:g6457790; PIDN:AAF09721.1; PID:g645779
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR0124
A:Map position: 1

Query Match 0.9%; Score 7; DB 2; Length 94;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 322 EMBEIRA 328
|||||
Db 81 EMBEIRA 87

RESULT 36

BVWYBA
Chaperonin groES - Mycobacterium tuberculosis
N:Alternate names: 10K antigen; BCG-a homolog; Cpn10; heat shock protein 10K (hsp10); im
C:Species: Mycobacterium tuberculosis
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jun-2000
C:Accession: S01381; A37166; S02727; A47292; G70737; A46481; S18040
R:Baird, P.N.; Hall, L.M.C.; Coates, A.R.M.
Nucleic Acids Res. 16, 9047, 1988
A:Title: A major antigen from Mycobacterium tuberculosis which is homologous to the heat
A:Reference number: S01381; MUID:89015584; PMID:2902558
A:Accession: S01381

A:Molecule type: DNA
 A:Residues: 1-100 <BA1>
 A:Cross-references: EMBL:X12598; NID:G44551; PID:G581358
 R.Baird, P.N.; Hall, L.M.C.; Coates, A.R.M.
 J. Gen. Microbiol. 135, 931-939, 1989
 A>Title: Cloning and sequence analysis of the 10 kDa antigen gene of *Mycobacterium tuberculosis*
 A:Reference number: A37166; NUID:90095443; PMID:2480990
 A:Accession: A37166
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 'V', 2-100 <BA2>
 A:Cross-references: GB:M25258; GB:X12598
 R.Shinnick, T.M.; Pilkey, B.B.; Hyche, A.D.; van Landingham, R.M.; Walker, L.L.
 Nucleic Acids Res. 17, 1254, 1989
 A>Title: The *Mycobacterium tuberculosis* BCG-a protein has homology with the *Escherichia*
 A:Reference number: S02727; NUID:89160258; PMID:2564178
 A:Accession: S02727
 A:Molecule type: DNA
 A:Residues: 1-100 <SH1>
 A:Cross-references: EMBL:X13739; NID:G44571; PID:CAA32003.1; PID:G581360
 R.Kong, T.H.; Coates, A.R.; Butcher, P.D.; Hickman, C.J.; Shinnick, T.M.
 Proc. Natl. Acad. Sci. U.S.A. 90, 2608-2612, 1993
 A>Title: *Mycobacterium tuberculosis* expresses two chaperonin-60 homologs.
 A:Reference number: A47292; NUID:93219332; PMID:7681982
 A:Accession: A47292
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-100 <KON>
 A:Cross-references: EMBL:X60350; NID:G44599; PID:CAA42908.1; PID:G581363
 A>Note: sequence extracted from NCBI backbone (NCBI:128605, NCBI:128607)
 R.Cole, S.T.; Brosch, R.; Parish, T.; Garnier, C.; Churcher, C.; Harris, D.; Gordon, S.
 & Connor, R.; Davies, R.; Parkhill, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellern, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrett, B.G.
 A>Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome
 A:Reference number: A70500; NUID:98295987; PMID:9634230
 A:Accession: G70737
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-100 <COL>
 A:Cross-references: GB:Z77165; GB:AL123456; NID:G3261609; PID:CA801005.1; PID:G1449369
 A:Experimental source: strain H37Rv
 R.Barnes, P.F.; Mehra, V.; Rivolta, B.; Fong, S.J.; Brennan, P.J.; Voegtline, M.S.; Mind
 J. Immunol. 148, 1835-1840, 1992
 A>Title: Immunoreactivity of a 10-kDa antigen of *Mycobacterium tuberculosis*.
 A:Reference number: A46481; NUID:92176646; PMID:1371791
 A:Accession: A46481
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 2-16 <BAR>
 A>Note: sequence extracted from NCBI backbone (NCBI:87128)
 C:Genetics:
 A:Gene: groES; cpn10
 A:Start codon: GTG
 C:Complex: functional chaperonin includes 14 chains of groEL and 7 of groES
 C:Function:
 A:Description: mediates protein folding and renaturation
 C:Superfamily: chaperonin groES
 C:Keywords: ATP; heat shock; molecular chaperone; stress-induced protein
 F:2-100/Product: chaperonin groES #status experimental <MAT>

Query Match 0.9%; Score 7; DB 1; Length 100;
 Best Local Similarity 100.0%; Pred. No. 62;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 605 GEXR1PL 611
 Db 55 GEXR1PL 61.

RESULT 37
 BWVY7B

chaperonin groES - *Mycobacterium bovis*
 N:Alternate names: heat shock protein, 10K (hsp10); immunogenic protein MPB57
 C:Species: *Mycobacterium bovis*
 C:Date: 31-Mar-1991 #sequence revision 31-Mar-1991 #ext_change 16-Jul-1999
 C:Accession: S01784; A37534; G60278
 R.Yamaguchi, R.; Matsuo, K.; Yamazaki, A.; Nagai, S.; Terasaka, K.; Yamada, T.
 FEBS Lett. 240, 115-117, 1988
 A>Title: Immunogenic protein MPB57 from *Mycobacterium bovis* BCG: molecular cloning, nuc
 A:Reference number: S01784; NUID:89052668; PMID:3056744
 A:Accession: S01784
 A:Molecule type: DNA
 A:Residues: 1-100 <YAM>
 A:Cross-references: EMBL:X13970; NID:G44190; PID:CAA32149.1; PID:G581313
 A:Accession: A37534
 A:Molecule type: protein
 A:Residues: 2-21 <YAM2>
 R.Filis, T.; Costopoulos, C.; Radford, A.J.; Bacic, A.; Wood, P.R.
 Infect. Immun. 59, 800-807, 1991
 A>Title: Purification and characterization of major antigens from a *Mycobacterium bovis*
 A:Reference number: A60278; NUID:91147217; PMID:1900061
 A:Accession: G60278
 A:Molecule type: protein
 A:Residues: 2-15, 'E', 17-20 <FIP>
 C:Genetics:
 A:Gene: groES
 A:Start codon: GTG
 C:Superfamily: chaperonin groES
 C:Keywords: heat shock; molecular chaperone; stress-induced protein
 F:2-100/Product: chaperonin groES #status predicted <MAT>

Query Match 0.9%; Score 7; DB 1; Length 100;
 Best Local Similarity 100.0%; Pred. No. 62;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 605 GEXR1PL 611
 Db 55 GEXR1PL 61

RESULT 38
 T38824
 thioltransferase - fission yeast (*Schizosaccharomyces pombe*)
 N:Alternate names: glutaredoxin
 C:Species: *Schizosaccharomyces pombe*
 C:Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #ext_change 20-Jun-2000
 C:Accession: T38824; T43362; T50489
 R.Connor, R.; Churcher, C.M.; Barrett, B.G.; Rajandream, M.A.; Wood, V.
 submitted to the EMBL Data Library, September 1997
 A:Reference number: Z21813
 A:Accession: T38824
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-101 <CON>
 A:Cross-references: EMBL:Z98980; PIDN:CA811722.1; GSPDB:GN00066; SPDB:SPAC4F10.20
 A:Experimental source: strain 97zh(-); cosmid c4F10
 R.Kawamukai, M.
 submitted to the EMBL Data Library, June 1998
 A:Description: S.pombe glutaredoxin.
 A:Reference number: Z22451
 A:Accession: T43362
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-101 <KAW>
 A:Cross-references: EMBL:AB015167; PIDN:BAA28750.1
 R.Kim, H.G.; Cho, Y.W.; Park, E.H.; Lim, C.J.
 submitted to the EMBL Data Library, January 1999
 A:Description: Characterization of cDNA encoding thioltransferase (glutaredoxin) from *S.
 A:Reference number: Z25091
 A:Accession: T50489
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-101 <KIM>
 A:Cross-references: EMBL:AF121275; PIDN:AAD25391.1*

C:Genetics:
A:Gene: grx1; SPAC4F10.20
A:Map position: 1
A:Int-cons: 7/3; 25/2; 58/1
C:Superfamily: glutaredoxin; glutaredoxin homology
C:Species: electron transfer; redox-active disulfide
C:Keywords: redox-active #status predicted
F:25-28/Disulfide bonds: redox-active #status predicted

Query Match 0.9%; Score 7; DB 2; Length 101;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 461 NSDFQAL 467
|||||
Db 82 NSDFQAL 88

RESULT 39
E82874
ribosomal protein S16 U568 [imported] - Ureaplasma urealyticum
C:Species: Ureaplasma urealyticum
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: E82874
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Helner, C.R.; Chen, E.Y.; Cassell, G.H.
submitted to GenBank, February 2000
A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a mir
A:Reference number: A82870
A:Accession: E82874
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-101 <STO>
A:Cross-references: GB:AE002155; GB:AF222894; NID:G5689572; PIDN:AAF30982.1; GSPDB:GN001
A:Experimental source: serovar 3, biovar 1
C:Genetics:
A:Gene: rps16; U568
A:Genetic code: SGC3

Query Match 0.9%; Score 7; DB 2; Length 101;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 404 PSETVKN 410
|||||
Db 66 PSETVKN 72

RESULT 40
I64171
hypothetical protein H1436 - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 09-Dec-2002
C:Accession: I64171
R:Feilschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kevlavage, A.
; Gozalayre, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodok, A.; Kelley, J.M.; Weidman, J.
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.
Science 269, 496-512, 1995
A:Authors: Ghem, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: I64171
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-106 <TIGR>
A:Cross-references: GB:U52822; GB:I42023; NID:g1574265; PIDN:AAC23085.1; PID:g1574275; T
A:Note: best homolog was a hypothetical protein from Erwinia carotovora
C:Superfamily: uncharacterized conserved protein

Query Match 0.9%; Score 7; DB 2; Length 106;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 207 AAEPFLS 213
|||||

Db 28 AAEPFLS 34

RESULT 41
D87498
HesB/YadR/YfhF family protein [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: D87498
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.O.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Koton
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: D87498
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-110 <STO>
A:Cross-references: GB:AE005673; NID:G13423480; PIDN:AAK23984.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC2009
C:Superfamily: conserved hypothetical protein H10376

Query Match 0.9%; Score 7; DB 2; Length 110;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 779 LALLKGS 785
|||||
Db 72 LALLKGS 78

RESULT 42
C86891
hypothetical protein ywab [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: C86891
R:Boletijn, A.; Wincker, P.; Manger, S.; Jallion, O.; Malarme, K.; Weissenbach, J.; Ehrlich,
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: C86891
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-110 <STO>
A:Cross-references: GB:AE005176; PID:g12725188; PIDN:AAK06229.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: ywab

Query Match 0.9%; Score 7; DB 2; Length 110;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 710 VEETPAE 716
|||||
Db 94 VEETPAE 100

RESULT 43
AG3598
transcription regulator [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
C:Accession: AG3598
R:DeVecchio, V.G.; Kaparat, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanova, I.
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AD3252; PMID:11756688
A:Accession: AG3598

A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-117 <RUR>
 A:Cross-references: GB:AE008918; PIDN:ALU53954.1; PID:G17984900; GSPDB:GN00191
 A:Experimental source: strain 16M
 C:Genetics:
 A:Gene: BME110712
 A:Map position: 11
 C:Superfamily: Rhizobium plasmid pNGR234a protein y4ho

Query Match 0.9%; Score 7; DB 2; Length 117;
 Best Local Similarity 100.0%; Pred. No. 71;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 LSASELA 206
 |||||
 Db 88 LSASELA 94

RESULT 44
 T44473

conserved hypothetical protein tnpG [imported] - Shigella flexneri
 C:Species: Shigella flexneri
 C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
 C:Accession: T44473

R:Moss, J.E.; Cardozo, T.J.; Zychlinsky, A.; Groisman, E.A.
 Mol. Microbiol. 33, 74-83, 1999
 A:Title: The selC-associated SHI-2 pathogenicity island of Shigella flexneri.
 A:Reference number: Z22779; MUID:99340540; PMID:10411725

A:Accession: T44473
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-121 <MOS>

A:Cross-references: EMBL:AF141323; NID:G5532445; PIDN:AD44740.1; PID:G5532456
 A:Experimental source: strain M90T; serotype 5a
 C:Genetics:
 A:Gene: tnpG

C:Superfamily: Agrobacterium tumefaciens insertion sequence IS1312 hypothetical protein

Query Match 0.9%; Score 7; DB 2; Length 121;
 Best Local Similarity 100.0%; Pred. No. 73;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 202 ASELAAA 208
 |||||
 Db 70 ASELAAA 76

RESULT 45
 NRPRH

pancreatic ribonuclease (EC 3.1.27.5) - pronghorn (tentative sequence)
 N:Alternate names: RNase 1; RNase A
 C:Species: Antilocapra americana (pronghorn)
 C:Date: 28-Feb-1981 #sequence_revision 28-Feb-1981 #text_change 31-Mar-2000
 C:Accession: A00813

R:Beintema, J.J.; Gastra, W.; Munnikema, J.
 J. Mol. Evol. 13, 305-316, 1979

A:Title: Primary structure of pronghorn pancreatic ribonuclease: close relationship betw
 A:Reference number: A00813; MUID:80075014; PMID:515141
 A:Accession: A00813

A:Molecule type: protein
 A:Residues: 1-124 <BBI>
 C:Superfamily: pancreatic ribonuclease
 C:Keywords: glycoprotein; hydrolase; nucleic acid digestion; pancreas
 F:12,41,119/Active site: His, Lys, His #status predicted
 F:26-84,40-95,58-110,65-72/Disulfide bonds: #status predicted
 F:34/Binding site: carbohydrate (asn) (covalent) (partial) #status experimental

Query Match 0.9%; Score 7; DB 1; Length 124;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 785 SNPSVS 791

Db 15 SNPSVS 21
 |||||

RESULT 46

hypothetical protein APE1982 - Aeropyrum pernix (strain K1)
 C:Species: Aeropyrum pernix
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
 C:Accession: H72500

R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Taka
 awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.;
 DNA Res. 6, 83-101, 1999

A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropy
 A:Reference number: A72450; MUID:99310339; PMID:10382966
 A:Accession: H72500

A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-128 <RAM>
 A:Cross-references: DBJ:AF000063; NID:G5105654; PIDN:BAA80992.1; PID:G5105680
 A:Experimental source: strain K1
 C:Genetics:
 A:Gene: APE1982

C:Superfamily: Aeropyrum pernix hypothetical protein APE1982

Query Match 0.9%; Score 7; DB 2; Length 128;
 Best Local Similarity 100.0%; Pred. No. 77;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 785 SNPSVS 791
 |||||
 Db 43 SNPSVS 49

RESULT 47
 T36220

mult-like protein - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 18-Aug-2000
 C:Accession: T36220

R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrett, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, March 1999
 A:Reference number: Z21577

A:Accession: T36220
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-129 <OLI>

A:Cross-references: EMBL:AL049573; PIDN:CA840318.1; GSPDB:GN00070; SCOEDB:SC39.10
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: SCOEDB:SC39.10

C:Superfamily: unassigned mult domain proteins; mult domain homology
 Query Match 0.9%; Score 7; DB 2; Length 129;
 Best Local Similarity 100.0%; Pred. No. 78;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 LTPDEV 39
 |||||
 Db 89 LTPDEV 95

RESULT 48

hypothetical protein slr0588 - Synechocystis sp. (strain PCC 6803)
 C:Species: Synechocystis sp.
 A:Variety: PCC 6803
 C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000

C:Accession: S76934
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.
 O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasud
 DNA Res. 3, 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocysti

5.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S76934
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-130 <KAN>
A:Cross-references: EMBL:D90917; GB:AB001339; NID:g1653836; PIDN:BA18846.1; PID:g165393
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: Synchocystis hypothetical protein slr0580

Query Match 0.9%; Score 7; DB 2; Length 130;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

268 DSLKQL 274
111 DSLKQL 117

RESULT 49
T35461
mutr-like protein - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 26-May-2000
C:Accession: T35461
R:Murphy, L.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1998
A:Reference number: Z21578
A:Accession: T35461
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-131 <MUR>
A:Cross-references: EMBL:AL031035; PIDN:CAA19915.1; GSPDB:GN00070; SCOPDB:SC6A9.30C
A:Genetics:
A:Experimental source: strain A312)
A:Superfamily: unassigned mutr domain proteins; mutr domain homology

Query Match 0.9%; Score 7; DB 2; Length 131;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

33 LTPDEV 39
91 LTPDEV 97

RESULT 50
B65240
hypothetical protein, 13.4K - Escherichia coli (strain K-12) insertion sequence IS2
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 23-Sep-1997 #text_change 01-Mar-2002
C:Accession: B65240; F64891; H64763; B65092; S56497; JQ0039
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burlingame, V.; Riley, M.;
A.; Rose, D.J.; Mau, B.; Mayo, S.; Fink, A.; Goehring, D.; Olsen, G.J.; White, O.;
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: B65240
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-136 <BLAT>
A:Cross-references: GB:AE000498; GB:U00096; NID:g2367366; PIDN:AACT7228.1; PID:g1790723;
A:Experimental source: strain K-12, substrain MGI655
A:Genetics: Y21
A:Accession: F64891
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-136 <BLAT>
A:Cross-references: GB:AE000237; GB:U00096; NID:g1787665; PIDN:AACT74485.1; PID:g1787670;
A:Experimental source: strain K-12, substrain MGI655
A:Genetics: Y22
A:Accession: H64763

A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-136 <BLAT>
A:Cross-references: GB:AE000493; GB:U00096; NID:g1786554; PIDN:AACT73463.1; PID:g1786557;
A:Experimental source: strain K-12, substrain MGI655
A:Genetics: Y11
A:Accession: B65092
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 16-136 <BLAT>
A:Cross-references: GB:AE000386; GB:U00096; NID:g2367187; PIDN:AACT6080.1; PID:g1789423;
A:Experimental source: strain K-12, substrain MGI655
A:Genetics: Y25
A:Accession: F65069
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 16-136 <BLAT>
A:Cross-references: GB:AE000369; GB:U00096; NID:g2367168; PIDN:AACT5900.1; PID:g1789225;
A:Experimental source: strain K-12, substrain MGI655
A:Genetics: Y24
R:Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.
Nucleic Acids Res. 23, 2105-2119, 1995
A:Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.
A:Reference number: S56314; MUID:95334362; PMID:7610040
A:Accession: S56497
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-136 <BUR>
A:Cross-references: EMBL:U14003; NID:g1263172; PIDN:AAA97168.1; PID:g537113
A:Genetics: Y26
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994
R:Ronecker, H.J.; Rak, B.
Gene 59, 291-296, 1987
A:Title: Genetic organization of insertion element IS2 based on a revised nucleotide seq
A:Reference number: A91582; MUID:88137965; PMID:2830172
A:Accession: JQ0039
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 16-136 <RON>
A:Cross-references: GB:V00279; GB:U01732; GB:M18426; NID:g41525; PIDN:CAA23542.1; PID:g4
A:Experimental source: K12, strain HB101
A:Genetics: HB1
C:Genetics: <Y21>
A:Gene: Y121
A:Molecule element: insertion sequence IS2
C:Genetics: <Y22>
A:Gene: Y121_2
A:Molecule element: insertion sequence IS2
C:Genetics: <Y11>
A:Gene: Y121_1
A:Molecule element: insertion sequence IS2
C:Genetics: <Y25>
A:Gene: Y121_5
A:Molecule element: insertion sequence IS2
C:Genetics: <Y24>
A:Gene: Y121_4
A:Molecule element: insertion sequence IS2
C:Genetics: <Y26>
A:Molecule element: insertion sequence IS2
C:Genetics: <HB1>
A:Molecule element: insertion sequence IS2
A:Start codon: GTG
C:Superfamily: Agrobacterium tumefaciens insertion sequence IS1312 hypothetical protein

Query Match 0.9%; Score 7; DB 2; Length 136;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

202 ASELAAA 208
85 ASELAAA 91

RESULT 51
D64964
hypothetical protein - Escherichia coli (strain K-12) insertion sequence IS2
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-sep-1997 #text_change 01-Mar-2002
C:Accession: D64964
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; PMID:97426617; PMID:92788503
A:Accession: D64964
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-136 <BLAT>
A:Cross-references: GB:AE000291; GB:U00096; NID:g1788298; PIDN:ACT5058.1; PID:g1788306;
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: y121_3
A:Mobile element: insertion sequence IS2
C:Superfamily: Agrobacterium tumefaciens insertion sequence IS1312 hypothetical protein

Query Match 0.9%; Score 7; DB 2; Length 136;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 202 ASELAAA 208
|||
|||
Db 85 ASELAAA 91

RESULT 52
T19240
hypothetical protein C13G3.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T19240
R:McMurray, A.
submitted to the EMBL Data Library, June 1996
A:Reference number: Z19096
A:Accession: T19240
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-136 <WILP>
A:Cross-references: EMBL:Z74027; PIDN:CAA98421.1; GSPDB:GN00023; CESP:C13G3.1
C:Genetics:
A:Gene: CESP:C13G3.1
A:Map position: 5
A:Introns: 49/1

Query Match 0.9%; Score 7; DB 2; Length 136;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 783 KGSNPS 789
|||
|||
Db 96 KGSNPS 102

RESULT 53
B81286
probable membrane protein Cj1410c [imported] - Campylobacter jejuni (strain NCTC 11168)
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C:Accession: B81286
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kettle, J.M.; Churcher, C.; Basham, D.; Chilling
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrell
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
A:Reference number: AB1250; PMID:20150912; PMID:10688204
A:Accession: B81286
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-139 <PAR>
A:Cross-references: GB:AL111168; GB:AL111168; NID:g6968723; PIDN:CAB73834.1; PID:g696881
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: Cj1410c
C:Superfamily: Campylobacter jejuni probable membrane protein Cj1410c

Query Match 0.9%; Score 7; DB 2; Length 139;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 IVIKITD 55
|||
|||
Db 100 IVIKITD 106

RESULT 54
A71717
hypothetical protein RP082 - Rickettsia prowazekii
C:Species: Rickettsia prowazekii
C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
C:Accession: A71717
R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark, T
Nature 396, 133-140, 1998
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A:Reference number: A71630; PMID:99039499; PMID:9823893
A:Accession: A71717
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-143 <AND>
A:Cross-references: GB:AJ235270; GB:AJ235269; NID:g3860572; PIDN:CA14552.1; PID:e134239
A:Experimental source: strain Madrid E
C:Genetics:
A:Gene: RP082

Query Match 0.9%; Score 7; DB 2; Length 143;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 NRKOEK 139
|||
|||
Db 106 NRKOEK 112

RESULT 55
AF1219
Salmonella enterica pduK protein homolog lmo1158 [imported] - Listeria monocytogenes (st
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AF1219
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
.J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entlian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karsic, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tietz, A.; Vazquez-Boland, U.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; PMID:21537279; PMID:11679669
A:Accession: AF1219
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-145 <GLA>
A:Cross-references: GB:NC_003210; PIDN:CAC99236.1; PID:g16410574; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo1158

Query Match 0.9%; Score 7; DB 2; Length 145;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 722 VETKVE 728

Db 103 VETEKVE 109

RESULT 56

Problematic transcription regulator PA3067 [imported] - *Pseudomonas aeruginosa* (strain PAO1)
 C:Species: *Pseudomonas aeruginosa*
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: D83261
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapidig, K.; Lim,
 ; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic patho
 A:Reference number: A82950; MUID:20437337; PMID:10984043
 A:Accession: D83261
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-147 <STO>
 A:Cross-references: GB:AE004731; GB:AE004091; NID:g9949171; PIDN:AA06455.1; GSPDB:GN001
 A:Experimental source: strain PAO1
 C:Genetics:
 A:Gene: PA3067
 C:Superfamily: Mycobacterium leprae hypothetical protein MLCB57.31

Query Match 0.9%; Score 7; DB 2; Length 147;
 Best Local Similarity 100.0%; Pred. No. 87;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 203 SELAAE 209
 Db 54 SELAAE 60

RESULT 57

Problematic transcription regulator PA3067 [imported] - *Escherichia coli* (strain O157:H
 C:Species: *Escherichia coli*
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: AB6079
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
 A:Reference number: AB6480; MUID:21074935; PMID:11206551
 A:Accession: AB6079
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-148 <STO>
 A:Cross-references: GB:AE005174; NID:g12518795; PIDN:AA059093.1; GSPDB:GN00145; UMGF:254
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: fliA
 C:Superfamily: phosphotransferase system enzyme II; phosphotransferase system mannitol-5

Query Match 0.9%; Score 7; DB 2; Length 148;
 Best Local Similarity 100.0%; Pred. No. 88;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 303 GVAAPHG 309
 Db 59 GVAAPHG 65

RESULT 58

Problematic transcription regulator PA3067 [imported] - *Escherichia coli* (strain O157
 C:Species: *Escherichia coli*
 C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 02-Nov-2001
 C:Accession: B91232
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and geno
 A:Reference number: A99629; MUID:21156231; PMID:11258796
 A:Accession: B91232
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-148 <HAV>
 A:Cross-references: GB:BA000007; PIDN:BA838249.1; PID:g13364302; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain RIMD 0509952
 C:Genetics:
 A:Gene: EC94826
 C:Superfamily: phosphotransferase system enzyme II; phosphotransferase system mannitol-5

Query Match 0.9%; Score 7; DB 2; Length 148;
 Best Local Similarity 100.0%; Pred. No. 88;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 303 GVAAPHG 309
 Db 59 GVAAPHG 65

RESULT 59

Problematic transcription regulator PA3067 [imported] - *Escherichia coli* (strain K-12)
 N:Alternate names: hypothetical protein f147 (rhad 3' region)
 C:Species: *Escherichia coli*
 C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 01-Mar-2002
 C:Accession: D48649; G65195; S40844; S36642
 R:Moralto, P.; Egan, S.M.; Hidalgo, E.; Aguilar, J.
 J. Bacteriol. 175, 5585-5594, 1993
 A:Title: Sequencing and characterization of a gene cluster encoding the enzymes for L-rh
 A:Reference number: A48649; MUID:93374854; PMID:8396120
 A:Accession: D48649
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-148 <MOR>
 A:Cross-references: GB:X60472; NID:g396678; PIDN:CAA43004.1; PID:g396682
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
 A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997.
 A:Title: The complete genome sequence of *Escherichia coli* K-12.
 A:Reference number: A64720; MUID:97426617; PMID:9278503
 A:Accession: G65195
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-148 <BLAT>
 A:Cross-references: GB:AE000465; GB:U00096; NID:g2367326; PIDN:AACT6882.1; PID:g2367327;
 A:Experimental source: strain K-12, substrain MG1655
 R:Plunkett III, G.; Burland, V.; Daniels, D.L.; Blattner, F.R.
 Nucleic Acids Res. 21, 3391-3398, 1993
 A:Title: Analysis of the *Escherichia coli* genome. III. DNA sequence of the region from 8
 A:Reference number: S40802; MUID:93347969; PMID:8346018
 A:Accession: S40844
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-80 'G', 82-102, 'XXXX', 107, 109-148 <BLU>
 A:Cross-references: EMBL:U13201; NID:g304961; PIDN:AA03033.1; PID:g305004
 A:Note: The nucleotide sequence was submitted to the EMBL Data Library, October 1993
 C:Genetics:
 A:Gene: fliA
 C:Superfamily: phosphotransferase system enzyme II; phosphotransferase system mannitol-5
 C:Keywords: phosphoprotein

Query Match 0.9%; Score 7; DB 2; Length 148;
 Best Local Similarity 100.0%; Pred. No. 88;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 303 GVAAPHG 309
 Db 59 GVAAPHG 65

```
RESULT 60
S39873
hypothetical protein 154 - Streptomyces ambofaciens
C:Species: Streptomyces ambofaciens
C>Date: 19-May-1994 #sequence_revision 26-May-1995 #text_change 26-May-2000
C:Accession: S39873; S33421
R:Hagege, J.; Perinot, J.; Friedmann, A.; Guerin, M.
Mol. Microbiol. 10, 799-812, 1993
A>Title: Mode and origin of replication of pSAM2, a conjugative integrating element of S
A:Reference number: S39873; PMID:95020551; PMID:7934842
A:Accession: S39873
A:Molecule type: DNA
A:Residues: 1-154 <HAG>
A:Cross-references: EMBL:Z19590; NID:G238044; PIDN:CAA79638.1; PID:G238045
A>Note: the authors translated the codon TAC for residue 81 as Thr and TAC for residue 1
C:Superfamily: unassigned mult domain proteins; mult domain homology
F:43-77/Domain: mult domain homology <MUT>

Query Match
Best Local Similarity 100.0%; Score 7; DB 2; Length 154;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 LTPDEVS 39
|||||
114 LTPDEVS 120

RESULT 61
AC2947
ribonuclease H [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AC2947
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.;
Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A>Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:21608550; PMID:11743193
A:Accession: AC2947
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-157 <KUR>
A:Cross-references: GB:AE008689; PIDN:AAL4393.1; PID:gl77441550; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
A:Genetics:
A:Gene: rnhA
A:Map position: linear chromosome
C:Superfamily: ribonuclease H

Query Match
Best Local Similarity 100.0%; Score 7; DB 2; Length 157;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 LDKLRL 473
|||||
117 LDKLRL 123

RESULT 62
G98335
ribonuclease H PA1815 [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C:Species: Agrobacterium tumefaciens
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C:Accession: G98335
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Oncollo, B.; Goldman,
A.; Liu, F.; Mollan, C.; Allinger, M.; Doughy, D.; Scott, C.; Leppas, C.; Markelz, B.;
Science 294, 2333-2328, 2001
A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; PMID:21608551; PMID:11743194
A:Accession: G98335
```

```
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-157 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK90209.1; PID:gl5160220; GSPDB:GN00170
A:Genetics:
A:Gene: AGR_L_3264
A:Map position: linear chromosome
C:Superfamily: ribonuclease H

Query Match
Best Local Similarity 100.0%; Score 7; DB 2; Length 157;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 LDKLRL 473
|||||
117 LDKLRL 123

RESULT 63
AF3240
conserved hypothetical protein Atu6104 [imported] - Agrobacterium tumefaciens (strain C
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AF3240
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.;
Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A>Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:21608550; PMID:11743193
A:Accession: AF3240
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-157 <KUR>
A:Cross-references: GB:AE008690; PIDN:AAL46340.1; PID:gl7744128; GSPDB:GN00189
A:Experimental source: strain C58 (Dupont)
A:Genetics:
A:Gene: Atu6104
A:Genome: plasmid

Query Match
Best Local Similarity 100.0%; Score 7; DB 2; Length 157;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 489 LAFAPL 495
|||||
5 LAFAPL 11

RESULT 64
D69256
hypothetical protein AF0052 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: D69256
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Usterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaite, B.P.; Sykes, S
Smith, H.O.; Woese, C.R.; Venter, J.C.
A>Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae
A:Reference number: A69250; PMID:98049343; PMID:938475
A:Accession: D69256
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-158 <LTF>
A:Cross-references: GB:AE001102; GB:AE000782; NID:G2689425; PIDN:AA91174.1; PID:G265055
C:Superfamily: Archaeoglobus fulgidus hypothetical protein AF2183

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 158;
```

Best Local Similarity 100.0%; Pred. No. 93;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 372 NSSLSVQ 378
|||||
Db 36 NSSLSVQ 42

RESULT 65

transcription elongation factor grea [imported] - Staphylococcus aureus (strain N315)

C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001

C:Accession: AB9943

R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguma, A.; Mizutani-U, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: AB9758; MUID:21311952; PMID:11418146

A:Accession: AB9943

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-158 <KUR>

A:Cross-references: GB:BA000018; PID:g13701408; PIDN:BAB42702.1; GSPDB:GN00149

A:Experimental source: strain N315

C:Genetics:

A:Gene: greA

C:Superfamily: transcription elongation factor greb

C:Keywords: transcription factor

Query Match 0.9%; Score 7; DB 2; Length 158;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 IIEDTGD 182
|||||
Db 77 IIEDTGD 83

RESULT 66

C83415

hypothetical protein PA1841 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: C83415

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim, J.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen.

A:Reference number: AB2950; MUID:20437337; PMID:10984043

A:Accession: C83415

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-165 <STO>

A:Cross-references: GB:AE004010; GB:AE004091; NID:g9947825; PIDN:AA05230.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA1841

Query Match 0.9%; Score 7; DB 2; Length 165;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 487 DLALFLA 493
|||||
Db 148 DLALFLA 154

RESULT 67

A86450

probable glycine cleavage system H-protein precursor - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Nov-2001

C:Accession: AB6450

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luos, J.S.; Malt, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Telson,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: AB6141; MUID:21016719; PMID:11130712

A:Accession: AB6450

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-166 <STO>

A:Cross-references: GB:AE005172; NID:g8920623; PIDN:AAF81345.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

C:Superfamily: glycine cleavage system protein H; lipoyl/biotin-binding homology

Query Match 0.9%; Score 7; DB 2; Length 166;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 788 SSVSKEK 794
|||||
Db 83 SSVSKEK 89

RESULT 68

C81235

transcription antitermination protein NusG NMB0126 [imported] - Neisseria meningitidis (

C:Species: Neisseria meningitidis

C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Feb-2001

C:Accession: C81235; AB2008

R:Terrellin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.

Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;

ti, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scariato, V.; Maignani, V.; Pizza, M.

Science 287, 1809-1815, 2000

A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve

A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.

A:Reference number: AB1000; MUID:20175755; PMID:10710307

A:Accession: C81235

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-178 <TER>

A:Cross-references: GB:AE002371; GB:AE002098; NID:g7225337; PIDN:AAF40585.1; PID:g722534

A:Experimental source: serogroup B, strain MC58

R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel

; Holtroyd, S.; Jørgensen, K.; Leather, S.; Moutle, S.; Mungall, K.; Quail, M.A.; Rajandream,

Nature 404, 502-506, 2000

A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491.

A:Reference number: AB1775; MUID:20222556; PMID:10761919

A:Accession: AB2008

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-178 <PAR>

A:Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CAB83462.1; PID:g737891

A:Experimental source: serogroup A, strain 22491

C:Genetics:

A:Gene: nusG; NMB0126; NMA0147

C:Superfamily: transcription antitermination factor nusG

Query Match 0.9%; Score 7; DB 2; Length 178;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 323 LEBRIAR 329
|||||
Db 22 LEBRIAR 28


```

RESULT 69
H71920
adenine phosphoribosyltransferase (EC 2.4.2.7) - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 23-Jul-1999
C:Accession: H71920
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
; Ives, C.; Gibson, R.; Metberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A:Reference number: A71800; MUID:99120557; PMID:9923682
A:Accession: H71920
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-179 <ARN>
A:Cross-references: GB:AE001485; GB:AE001439; NID:g4155057; PIDN:AAD06100.1; PID:g415506
A:Experimental source: strain J99
C:Genetics:
A:Gene: apt
C:Function:
A:Pathway: purine salvage
C:Superfamily: adenine phosphoribosyltransferase
C:Keywords: glycosyltransferase; pentosyltransferase; salvage pathway

Query Match
Best Local Similarity 0.9%; Score 7; DB 2; Length 179;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 484 LVDDLLA 490
|||||
DB 124 LVDDLLA 130

RESULT 70
A70178
conserved hypothetical protein B80626 - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 09-Dec-2002
C:Accession: A70178
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Jathigra, R.; White
son, D.; Peterson, J.; Kexlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943; PMID:9403685
A:Accession: A70178
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-181 <RLE>
A:Cross-references: GB:AE001164; GB:AE000783; NID:g2688550; PIDN:AAC66987.1; PID:g268855
A:Experimental source: strain B31
C:Superfamily: Primase-related protein

Query Match
Best Local Similarity 0.9%; Score 7; DB 2; Length 181;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 469 KLLERLN 475
|||||
DB 153 KLLERLN 159

RESULT 71
F75487
v-type ATP synthase, E subunit - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: F75487
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

```

```

, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioreistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: F75487
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-185 <WHI>
A:Cross-references: GB:AE001926; GB:AE000513; NID:g6458395; PIDN:AAF10275.1; PID:g645840
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR0697
A:Map position: 1

Query Match
Best Local Similarity 0.9%; Score 7; DB 2; Length 185;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 466 ALDKLLE 472
|||||
DB 2 ALDKLLE 8

RESULT 72
T35792
beta-glucosidase - Streptomyces coelicolor (fragment)
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 26-May-2000
C:Accession: T35792
R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1998
A:Reference number: Z21570
A:Accession: T35792
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-191 <SEB>
A:Cross-references: EMBL:AL031013; PIDN:CAA19797.1; GSPDB:GN000070; SCOEDB:SC8A6.25
C:Genetics:
A:Experimental source: strain A312)
C:Genetics:
A:Gene: SCOEDB:SC8A6.25
C:Superfamily: Agrobacterium beta-glucosidase

Query Match
Best Local Similarity 0.9%; Score 7; DB 2; Length 191;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 484 LVDDLLA 490
|||||
DB 122 LVDDLLA 128

RESULT 73
S01249
cytochrome c-L precursor - Methylobacterium sp.
C:Species: Methylobacterium sp.
C:Date: 30-Sep-1989 #sequence_revision 02-Jul-1996 #text_change 21-Jul-2000
C:Accession: S01249; S02658
R:Nunn, D.N.; Anthony, C.
Biochem. J. 256, 673-676, 1988
A:Title: The nucleotide sequence and deduced amino acid sequence of the cytochrome c(L)
A:Reference number: S02658; MUID:89134152; PMID:2851998
A:Accession: S02658
A:Molecule type: DNA
A:Residues: 1-197 <NUN2>

```

A>Note: source designated as *Methylobacterium extorquens* AM1
A>Note: part of this sequence was confirmed by protein sequencing
C:Genetics:
A:Gene: moxG
C:Function:
A:Description: electron acceptor for methanol dehydrogenase; electron donor to cytochrom
A:Pathway: methanol oxidation
C:Superfamily: cytochrome c-L; cytochrome c6 homology
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; periplasmic sp
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-197/Product: cytochrome c-L #status predicted <MAT>
F:80-153/Domain: cytochrome c6 homology <CY6>
F:90-93/Binding site: heme (Cys) (covalent) #status predicted
F:94-134/Binding site: heme iron (His, Met) (axial ligands) #status predicted

Query Match 0.9%; Score 7; DB 1; Length 197;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 ENLTPDE 37
|||||
DB 139 ENLTPDE 145

RESULT 74
S06147
GTP-binding protein rab1B - rat
N:Alternate names: rat-related protein rab1B
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 02-Feb-2001
C:Accession: S06147; S03189
R:Touchot, N.; Zahraoui, A.; Vieh, E.; Tavittian, A.
FEBS Lett. 256, 79-84, 1989
A:Title: Biochemical properties of the YPT-related rab1 protein. Comparison with rab1A.
A:Reference number: S06147; MUID:90033316; PMID:2509243
A:Accession: S06147
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-201 <TOU>
R:Vieh, E.; Touchot, N.; Zahraoui, A.; Tavittian, A.
Nucleic Acids Res. 17, 1770, 1989
A:Title: Nucleotide sequence of a rat cDNA, RAB1B, encoding a RAB1-YPT related protein.
A:Reference number: S03189; MUID:89160341; PMID:2493636
A:Accession: S03189
A:Molecule type: mRNA
A:Residues: 1-95, 'YA', 98-201 <VIE>
A:Cross-references: EMBL:X13905; NID:G57005; PIDN:CAA32105.1; PID:G57006
C:Genetics:
A:Gene: rab1B
C:Superfamily: ras transforming protein; translation elongation factor Tu homology
C:Keywords: GTP binding; lipoprotein; membrane trafficking; nucleotide binding; P-loop;
F:9-124/Domain: translation elongation factor Tu homology <ETU>
F:15-22/Region: nucleotide-binding motif A (P-loop)
F:121-124/Region: GTP-binding NKXD motif
F:151-153/Region: GTP-binding SAK/L motif
F:21-22, 40, 121, 122, 124, 151/Binding site: Mg-GTP (Lys, Ser, Thr, Asn, Lys, Asp, Ser) #sta
F:200, 201/Binding site: geranyl-geranyl (Cys) (covalent) #status predicted

Query Match 0.9%; Score 7; DB 2; Length 201;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 365 PNLKIDS 371
|||||
DB 184 PNLKIDS 190

RESULT 75
P90658
hypothetical protein EC60238 [imported] - *Escherichia coli* (strain O157:H7, substrain RI
C:Species: *Escherichia coli*
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: P90658

R.Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
Gasawara, N.; Yasunaga, T.; Kuhnara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and geno
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: P90658
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-203 <HAV>
A:Cross-references: GB:BA000007; PIDN:BA033661.1; PID:G13359694; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain R1MD 0509952
C:Genetics:
A:Gene: EC60238

Query Match 0.9%; Score 7; DB 2; Length 203;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 774 EAEKLLA 780
|||||
DB 31 EAEKLLA 37

RESULT 76
P85509
hypothetical protein Z0269 [imported] - *Escherichia coli* (strain O157:H7, substrain EDL9
C:Species: *Escherichia coli*
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: P85509
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: P85509
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-203 <STO>
A:Cross-references: GB:AE005174; NID:G12512978; PIDN:AG54538.1; GSPDB:GN00145; UWGP:202
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z0269

Query Match 0.9%; Score 7; DB 2; Length 203;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 774 EAEKLLA 780
|||||
DB 31 EAEKLLA 37

RESULT 77
S62443
diadenosine 5', 5''-P1, P6-hexaphosphate hydrolase [validated] - fission yeast (*Schizosa
N:Alternate names: hypothetical protein SPAC24B11.03
C:Species: *Schizosaccharomyces pombe*
C:Date: 16-May-1996 #sequence_revision 13-Mar-1997 #text_change 15-Sep-2000
C:Accession: S62443; S62548; T37649; T38330; T50491
R:Odell, C.; Bowman, S.
submitted to the EMBL Data Library, October 1995
A:Reference number: S62430
A:Accession: S62443
A:Molecule type: DNA
A:Residues: 1-210 <ODE>
A:Cross-references: EMBL:Z54308; NID:G1008985; PIDN:CAA91107.1; PID:G1008999
R:Odell, C.; Churcher, C.M.
submitted to the EMBL Data Library, November 1995
A:Reference number: S62546
A:Accession: S62548
A:Molecule type: DNA
A:Residues: 1-210 <OD2>
A:Cross-references: EMBL:Z67757; NID:G1061288; PIDN:CAA91768.1; PID:G1061291*

R.Odeall, C.; Bowman, S.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.; Wood, V.
submitted to the EMBL Data Library, October 1995
A:Reference number: Z21734
A:Accession: T37649
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-210 <OD3>
A:Cross-references: EMBL:Z54308; PIDN:CAA91107.1; GSPDB:GN00066; SPDB:SPAC24B11.14
A:Experimental source: strain 972h(-); cosmid c1366
R.Odeall, C.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, November 1995
A:Reference number: Z21786
A:Accession: T38330
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-210 <OD4>
A:Cross-references: EMBL:Z67757; PIDN:CAA91768.1; GSPDB:GN00066; SPDB:SPAC24B11.03
A:Experimental source: strain 972h(-); cosmid c24B11
R.Ingram, S.W.; Stratemann, S.A.; Barnes, L.D.
Biochemistry 38, 3649-3355, 1999
A:Title: Schizosaccharomyces pombe Aps1, a diadenosine 5',5''-P1,P6-hexaphosphate hydrolase purified enzyme.
A:Reference number: Z25093; MID:99192402; PMID:10090752
A:Accession: T50491
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-210 <ING>
A:Cross-references: EMBL:AF15215; PIDN:AAD2005.1
A:Experimental source: strain 972
C:Genetics:
A:Gene: aps1; SPAC13G6.14; SPAC24B11.03
A:Map position: 1L
C:Complex: monomer [validated, MUID:99192402]
C:Function:
A:Description: catalyzes the hydrolysis of dinucleoside oligophosphates [validated, MUID:99192402]
C:Superfamily: Saccharomycetes cerevisiae hypothetical protein YOR163w

Query Match
Best Local Similarity 100.0%; Score 7; DB 2; Length 210;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 773 AEAEXLL 779
|||||
Db 143 AEAEXLL 149

RESULT 78
AH0899
conserved hypothetical protein STY3451 [imported] - Salmonella enterica subsp. enterica
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AH0899
R.Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Garra, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AH0899
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-211 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD07790.1; PID:G16504338; GSPDB:GN00176
C:Genetics:
A:Gene: STY3451

Query Match
Best Local Similarity 100.0%; Score 7; DB 2; Length 211;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 600 YNRVKG 606

Db 122 YNRVKG 128
|||||

RESULT 79
T01627
Probable ATP binding protein At2g18990 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 02-Feb-2001
C:Accession: T01627; B84571
R.Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
submitted to the EMBL Data Library, April 1998
A:Description: Arabidopsis thaliana chromosome II BAC F19F24 genomic sequence.
A:Reference number: Z14153
A:Accession: T01627
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-211 <ROU>
A:Cross-references: EMBL:AC003673; MID:G3004543; PID:G3004560
A:Experimental source: cultivar Columbia
R.Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M. Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L. eues, D.; Niernan, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Frazer, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: AB4420; MUID:20083487; PMID:1061197
A:Accession: B84571
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-211 <STO>
A:Cross-references: GB:AE002093; MID:G6598372; PIDN:AAF18603.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g18990
A:Map position: 2
A:Insertions: 7/3; 99/3; 152/3
A:Note: F19F24.19

Query Match
Best Local Similarity 100.0%; Score 7; DB 2; Length 211;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 323 LERIRAR 329
|||||
Db 169 LERIRAR 175

RESULT 80
F83144
conserved hypothetical protein PA4017 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: F83144
R.Stover, C.K.; Phan, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lartig, K.; Lam, J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
A:Reference number: AB2950; MUID:20437337; PMID:10984043
A:Accession: F83144
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-213 <STO>
A:Cross-references: GB:AE004818; GB:AE004091; MID:G9950200; PIDN:AA07404.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA4017

Query Match
Best Local Similarity 100.0%; Score 7; DB 2; Length 213;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 600 YNRVKG 606
|||||

Db 125 YNRVKGE 131

RESULT 81

hypothetical protein - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 28-Jul-2000

C:Accession: B75533

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uetebach, T.; Zalewski, C.; Ma S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896; PMID:10567266

A:Accession: B75533

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-219 <WHI>

A:Cross-references: GB:AE001893; GB:AE00513; NID:g6458001; PIDN:AAF0916.1; PID:g645801

A:Experimental source: strain R1

A:Genetics:

A:Gene: DR0333

A:Map position: 1

C:Superfamily: Deinococcus radiodurans hypothetical protein DR0333

Query Match 0.9%; Score 7; DB 2; Length 219;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 199 ELSASEL 205

Db 105 ELSASEL 111

RESULT 82

DE5105

hypothetical 24.8 kD protein in agaI-mtr intergenic region - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002

C:Accession: D65105

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A61720; MUID:97426617; PMID:9278503

A:Accession: D65105

A:Status: Preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-226 <BLAT>

A:Cross-references: GB:AE000396; GB:U00096; NID:g1789536; PIDN:AACT6186.1; PID:g1789542;

A:Experimental source: strain K-12, substrain MG1655

A:Genetics:

A:Gene: yrar

Query Match 0.9%; Score 7; DB 2; Length 226;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 600 YNRVKGE 606

Db 137 YNRVKGE 143

RESULT 83

A98133

hypothetical protein Ecs4033 [imported] - Escherichia coli (strain O157:H7, substrain R1)

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C:Accession: A98133

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Gasevarea, N.; Yasunaga, T.; Kuhara, S.; Shibata, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno

A:Reference number: A98629; MUID:21156231; PMID:11258796

A:Accession: A98133

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-226 <HAV>

A:Cross-references: GB:BA000007; PIDN:BA37456.1; PID:g13363506; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain R1MD 0509952

C:Genetics:

A:Gene: Ecs4033

Query Match 0.9%; Score 7; DB 2; Length 226;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 600 YNRVKGE 606

Db 137 YNRVKGE 143

RESULT 84

DE5978

hypothetical protein yrar [imported] - Escherichia coli (strain O157:H7, substrain EDL93)

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C:Accession: D85978

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalante, E.; Potamousis, K.; Apodaca,

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: D85978

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-226 <STO>

A:Cross-references: GB:AE005174; NID:g12517760; PIDN:AAG58288.1; GSPDB:GN00145; UMG:Z45

A:Experimental source: strain O157:H7, substrain EDL933

A:Genetics:

A:Gene: yrar

Query Match 0.9%; Score 7; DB 2; Length 226;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 600 YNRVKGE 606

Db 137 YNRVKGE 143

RESULT 85

GB1894

probable integral membrane protein NMA1265 [imported] - Neisseria meningitidis (strain Z)

C:Species: Neisseria meningitidis

C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001

C:Accession: G81894

R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel

; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,

Nature 404, 502-506, 2000

A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z491.

A:Reference number: A81775; MUID:20222556; PMID:10761919

A:Accession: G81894

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-226 <PAR>

A:Cross-references: GB:AL162755; GB:AL157959; NID:g7379742; PIDN:CAB84520.1; PID:g737994

A:Experimental source: serogroup A, strain Z491

A:Genetics:

A:Gene: NMA1265

Query Match 0.9%; Score 7; DB 2; Length 226;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 AEAFLSG 214
 |||||
 Db 189 AEAFLSG 195

RESULT 86
 DB1125
 hypothetical protein NMB1066 [imported] - Neisseria meningitidis (strain MC58 serogroup C)
 C/Species: Neisseria meningitidis
 C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
 C/Accession: DB1125
 R/Retcelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masiatani, V.; Pizza, M. Science 287, 1809-1815, 2000
 A/Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappolli, R.; V. A/Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
 A/Reference number: AB1000; MUID:20175755; PMID:10710307
 A/Accession: DB1125
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-226 <TET>
 A/Cross-references: GB:AE002457; GB:AE002099; NID:g7226304; PIDN:AAFA41462.1; PID:g722630
 A/Experimental source: serogroup B, strain MC58
 C/Genetics:
 A/Gene: NMB1066

Query Match 0.9%; Score 7; DB 2; Length 226;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 AEAFLSG 214
 |||||
 Db 189 AEAFLSG 195

RESULT 87
 CB3744
 two-component response regulator BH0755 [imported] - Bacillus halodurans (strain C-125)
 C/Species: Bacillus halodurans
 C/Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
 C/Accession: CB3744
 R/Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira Nucleic Acids Res. 28, 4317-4331, 2000
 A/Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and A/Reference number: AB3650; MUID:20512582; PMID:11058132
 A/Accession: CB3744
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-232 <STO>
 A/Cross-references: GB:AP001509; GB:BA000004; NID:g10173176; PIDN:BA04474.1; GSPDB:GN00
 A/Experimental source: strain C-125
 C/Genetics:
 A/Gene: BH0755
 C/Superfamily: ompr protein; response regulator homology

Query Match 0.9%; Score 7; DB 2; Length 232;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 735 EVLLAKV 741
 |||||
 Db 106 EVLLAKV 112

RESULT 88
 G87286
 conserved hypothetical protein CC0304 [imported] - Caulobacter crescentus
 C/Species: Caulobacter crescentus
 C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
 C/Accession: G87286
 R/Nierman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J. B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A/Title: Complete Genome Sequence of Caulobacter crescentus.
 A/Reference number: AB7249; MUID:21173698; PMID:1125647
 A/Accession: G87286
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-237 <STO>
 A/Cross-references: GB:AE005673; NID:g13421447; PIDN:AAK22291.1; GSPDB:GN00148
 C/Genetics:
 A/Gene: CC0304

Query Match 0.9%; Score 7; DB 2; Length 237;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 GAVALLR 158
 |||||
 Db 152 GAVALLR 158

RESULT 89
 I67638
 proteasome activator rPA28 subunit beta - rat
 C/Species: Rattus sp. (rat)
 C/Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 24-Sep-1999
 C/Accession: I67638
 R/Joan Young, A.; Tanahashi, N.; Akiyama, K. FEBS-Lett. 366, 37-42, 1995
 A/Title: Primary structures of two homologous subunits of PA28, a gamma-interferon-induc A/Reference number: I53518; MUID:95309399; PMID:7789512
 A/Accession: I67638
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-238 <RES>
 A/Cross-references: GB:DA5250; NID:g1008931; PIDN:BA08207.1; PID:g1008932
 C/Superfamily: proteasome activator PA28 alpha chain

Query Match 0.9%; Score 7; DB 2; Length 238;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 776 EKLLALL 782
 |||||
 Db 97 EKLLALL 103

RESULT 90
 D75203
 hypothetical protein PAB2251 - Pyrococcus abyssi (strain Orsay)
 C/Species: Pyrococcus abyssi
 C/Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
 C/Accession: D75203
 R/anonymous, Genoscope submitted to the EMBL Data Library, July 1999
 A/Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru A/Reference number: A75001
 A/Accession: D75203
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-240 <KAW>
 A/Cross-references: GB:AJ248283; GB:AL096836; NID:g5457433; PIDN:CAB49075.1; PID:e151496
 A/Experimental source: strain Orsay
 C/Genetics:
 A/Gene: PAB2251

Query Match 0.9%; Score 7; DB 2; Length 240;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 LKEAEVL 737
 |||||
 Db 96 LKEAEVL 102

```
RESULT 91
C71725
amino-acid ABC transporter binding protein (yqix) RP150 - Rickettsia prowazekii
C:Species: Rickettsia prowazekii
C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
C:Accession: C71725
R:Anderson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichertz-Fonten, T.; Alsmark, U.
Nature 396, 133-140, 1998
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A:Reference number: A71630; MUID:99039499; PMID:9823893
A:Accession: C71725
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-241 <AND>
A:Cross-references: GB:AJ235270, GB:AJ235269, NID:g3860572, PIDN:CAA14618.1, PID:g386071
A:Experimental source: strain Madrid E
C:Genetics:
A:Gene: yqix; RP150
C:Superfamily: lysine-arginine-ornithine-binding protein

Query Match          0.9%; Score 7; DB 2; Length 241;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 76 DALISEE 82
Db 173 DALISEE 179

RESULT 92
H97723
hypotheical protein yqix [imported] - Rickettsia conorii (strain Malish 7)
C:Species: Rickettsia conorii
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 02-Nov-2001
C:Accession: H97723
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro
Science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MUID:21442074; PMID:11557893
A:Accession: H97723
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-241 <KID>
A:Cross-references: GB:AE006914; PIDN:AAL02730.1; PID:g15619241; GSPDB:GN00173
C:Genetics:
A:Gene: yqix
C:Superfamily: lysine-arginine-ornithine-binding protein

Query Match          0.9%; Score 7; DB 2; Length 241;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 76 DALISEE 82
Db 173 DALISEE 179

RESULT 93
T03482
iron(III) dicitrate transporter ATP-binding protein - Rhodobacter capsulatus
C:Species: Rhodobacter capsulatus
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 08-Oct-1999
C:Accession: T03482
R:Vilek, C.; Paces, V.; Maltsev, N.; Paces, J.; Haselkorn, R.; Fomstein, M.
Proc. Natl. Acad. Sci. U.S.A. 94, 9384-9388, 1997
A:Title: Sequence of a 189-kb segment of the chromosome of Rhodobacter capsulatus SB1003
A:Reference number: Z14955; MUID:97404404; PMID:9256491
A:Accession: T03482
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-255 <VLC>
```

```
A:Cross-references: EMBL:AF010496; NID:g3128256; PIDN:AA016135.1; PID:g3128283
C:Genetics:
A:Gene: fecC
A:Map position: 1
C:Superfamily: inner membrane protein malK; ATP-binding cassette homology
C:Keywords: iron transport

Query Match          0.9%; Score 7; DB 2; Length 255;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 778 L1ALLG 784
Db 181 L1ALLG 187

RESULT 94
I38025
keratin-like protein - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 13-Aug-1999
C:Accession: I38025; S60680
R:Tomasetto, C.; Regnier, C.H.; Moog-Lutz, C.; Mattei, M.G.; Chenard, M.P.; Lidereau, R.
Genomics 28, 367-376, 1995
A:Title: Identification of four novel human genes amplified and overexpressed in breast
A:Reference number: I37080; MUID:96039245; PMID:7490069
A:Accession: I38025
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-257 <RES>
A:Cross-references: EMBL:X80197; NID:g951271; PIDN:CAA56488.1; PID:g951272
C:Genetics:
A:Gene: MN137
C:Superfamily: cytoskeletal keratin

Query Match          0.9%; Score 7; DB 2; Length 257;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 127 RTKEIN 133
Db 69 RTKEIN 75

RESULT 95
D86722
hypotheical protein accA [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 17-May-2002
C:Accession: D86722
R:Boletijn, A.; Wincker, P.; Manger, S.; Jallou, O.; Malame, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: D86722
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-260 <STO>
A:Cross-references: GB:AE005176; PID:g12723699; PIDN:AAK04878.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: accA
C:Superfamily: acetyl-CoA carboxylase, carboxyltransferase alpha chain

Query Match          0.9%; Score 7; DB 2; Length 260;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 76 DALISEE 82
Db 222 DALISEE 228
```

RESULT 96
G70980
transcription initiation factor sigma F - Mycobacterium tuberculosis (strain H37Rv)
N:Alternate names: sigF protein; sporulation sigma factor homolog; stress-response sigma
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: G70980; J06131
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Feltywell, T.; Genies, S.; Hamlin, N.; Holroyd, S.
; Rajandream, M.A.; Rogers, J.; Ruter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
R:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: G70980
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-261 <COL>
A:Cross-references: GB:292771; GB:AL123456; NID:93242259; PIDN:CAB07069.1; PID:G1877341
A:Experimental source: strain H37Rv
R:Demajo, J.; Zhang, Y.; Ko, C.; Young, D.B.; Bishai, W.R.
Proc. Natl. Acad. Sci. U.S.A. 93, 2790-2794, 1996
A:Title: A stationary-phase stress-response sigma factor from Mycobacterium tuberculosis
A:Reference number: J06131; MUID:96181544; PMID:8610119
A:Accession: J06131
A:Molecule type: DNA
A:Residues: 1-247, 'V', 249-261 <DEM>
A:Cross-references: GB:U41061; NID:91276888; PIDN:AA04103.1; PID:G1276889
C:Comment: This factor is a subunit of bacterial RNA polymerase and confer promoter spec
and persist during human infection.
C:Genetics:
A:Gene: sigF
A:Start codon: GTG
C:Superfamily: transcription sigma factor G; transcription initiation factor sigma katF
C:Keywords: DNA binding; sigma factor; transcription initiation
F:45-261/Domain: transcription initiation factor sigma katF homology <KTF>
Query Match
Best Local Similarity 0.9%; Score 7; DB 2; Length 261;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 201 SASLEAA 207
Db 143 SASLEAA 149
RESULT 97
D71037
hypothetical protein PH1588 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
C:Accession: D71037
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Ohikku, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushiida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A:Reference number: A71000; MUID:98344137; PMID:9679194
A:Accession: D71037
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-264 <KAW>
A:Cross-references: GB:AP000006; NID:93236133; PIDN:BA030700.1; PID:G3256017
A:Experimental source: strain OT3
A>Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH1588
C:Superfamily: heat-shock protein htpX
Query Match
Best Local Similarity 0.9%; Score 7; DB 2; Length 264;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 325 ERIARI 331

Db 245 ERIARI 251
RESULT 98
T32639
hypothetical protein F42G8.11 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T32639
R:Gatung, S.; Holmes, A.
submitted to the EMBL Data Library, December 1997
A:Description: The sequence of C. elegans cosmid F42G8.
A:Reference number: 221203
A:Accession: T32639
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-269 <GAT>
A:Cross-references: EMBL:AF038618; PIDN:AA092070.1; GSPDB:GN00022; CESP:F42G8.11
A:Experimental source: strain Bristol NZ; clone F42G8
C:Genetics:
A:Gene: CESP:F42G8.11
A:Map position: 4
A:Introns: 40/1; 67/3; 107/2; 236/3
Query Match
Best Local Similarity 0.9%; Score 7; DB 2; Length 269;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 606 EKRIPLV 612
Db 165 EKRIPLV 171
RESULT 99
E0RT2
pancreatic elastase II (EC 3.4.21.71) precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 18-Aug-1992 #sequence_revision 18-Aug-1992 #text_change 24-Sep-1999
C:Accession: A00961
R:Macdonald, R.J.; Swift, G.H.; Quinto, C.; Swain, W.; Pictet, R.L.; Nikovits, W.; Rutte
Biochemistry 21, 1453-1463, 1982
A:Title: Primary structure of two distinct rat pancreatic preproelastases determined by
A:Reference number: A00960; MUID:82182967; PMID:6918221
A:Accession: A00961
A:Molecule type: mRNA
A:Residues: 1-271 <MAC>
A:Cross-references: GB:L00124; GB:J00731; NID:9204019; PIDN:AAA98780.1; PID:G204021
C:Superfamily: trypsin; trypsin: homology
C:Keywords: hydrolase; pancreas; serine proteinase; zymogen
F:11-16/Domain: signal sequence #status predicted <SIG>
F:17-30/Domain: activation peptide #status predicted <APT>
F:31-271/Product: elastase II #status predicted <MP>
F:31-264/Domain: trypsin homology <TRY>
F:75,123,218/Active site: His, Asp, Ser #status predicted
Query Match
Best Local Similarity 0.9%; Score 7; DB 1; Length 271;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 219 SNSRYR 225
Db 78 SNSRYR 84
RESULT 100
S62364
L-2-chloropropionic acid dehalogenase (EC 3.8.1.-) - Rhizobium sp.
N:Alternate names: L-2-haloalkanoate dehalogenase
C:Species: Rhizobium sp.
C>Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000
C:Accession: S62364
R:Cairns, S.S.; Cornish, A.; Cooper, R.A.

Eur. J. Biochem. 235, 744-749, 1996
A:Title: Cloning, sequencing and expression in Escherichia coli of two Rhizobium sp. gen
A:Reference number: S62363; MUID:96184554; PMID:8654424
A:Accession: S62364
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-279 <CAI>
A:Cross-references: EMBL:X93597; NID:G1103493; PIDN:CA63794.1; PID:G1103495
C:Superfamily: Rhizobium L-2-chloropropionic acid dehalogenase
C:Keywords: hydrolase

Query Match 0.9%; Score 7; DB 2; Length 279;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 744 SSLKANA 750
|||
Db 243 SSLKANA 249

Search completed: November 14, 2003, 11:02:04
Job time : 56 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 10:44:30 ; Search time 24 Seconds
(without alignments)
1559.719 Million cell updates/sec

Title: US-09-765-271-56

Perfect score: 796

Sequence: 1 SYELGLYQARTVKNRVS.....KLALIKGSNPSVSKKIN 796

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 200 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	1.1	921	1 T2D3 DROME	P47825 drosophila
2	8	1.0	71	1 Y029_NPVAC	P41433 autographa
3	8	1.0	292	1 SCI7_NEUCR	O9665 neurospora
4	8	1.0	448	1 AAC2_DICDI	P14196 dictyostell
5	8	1.0	503	1 IRF7_HUMAN	O92985 homo sapien
6	8	1.0	795	1 LON_MYCN	P78025 mycoplasma
7	8	1.0	861	1 Y240_ARATH	O22224 arabidopsis
8	8	1.0	1080	1 HDC_DROME	O9n2m8 drosophila
9	8	1.0	1205	1 SMC_MYCTU	O92817 homo sapien
10	8	1.0	2033	1 EVPL_HUMAN	O9d952 mus musculu
11	8	1.0	2035	1 EVPL_MOUSE	O9d952 mus musculu
12	7	0.9	70	1 Y872_PIRAB	O9v0b8 pyrococcus
13	7	0.9	99	1 CH10_MYCBO	P15020 mycobacteri
14	7	0.9	99	1 CH10_MYCBO	P15020 mycobacteri
15	7	0.9	101	1 GLR1_SCHPO	O36032 schizosacch
16	7	0.9	101	1 RS16_UREPA	O9p9s1 ureaplasma
17	7	0.9	106	1 Y0CC_HAEIN	O57152 haemophilus
18	7	0.9	115	1 RNPA_STRAP	O8cmu4 staphylococ
19	7	0.9	121	1 INSC_ECOLI	P19776 escherichia
20	7	0.9	121	1 RNP_ANTAM	P59444 shigella fl
21	7	0.9	124	1 INSC_SHIFL	P00668 antillocapra
22	7	0.9	143	1 Y082_RICPR	O92655 rickettsia
23	7	0.9	148	1 PTVA_ECOLI	P32155 escherichia
24	7	0.9	154	1 MUTT_STRAM	P32031 streptomyc
25	7	0.9	158	1 GREB_STRAM	O99tn9 staphylococ
26	7	0.9	166	1 Y052_ARCFU	O9184 archaeglob
27	7	0.9	174	1 GCSI_ARATH	O91q10 arabidopsis
28	7	0.9	174	1 HUNB_DROMU	O46250 drosophila
29	7	0.9	177	1 APT_CHUTE	O8kfm9 chlorobium
30	7	0.9	178	1 NUSG_NEIMA	O9j1d9 neisseria m
31	7	0.9	179	1 APT_HELPU	O9z1g9 helicobacte
32	7	0.9	185	1 VATE_DEIRA	O9rwl1 deinococcus
33	7	0.9	197	1 CYCL_METEX	P14774 methyllobact

34	7	0.9	201	1 RB1B_HUMAN	O9hou4 homo sapien
35	7	0.9	201	1 RB1B_RAT	P10536 rattus norv
36	7	0.9	210	1 AP51_SCHPO	O09790 schizosacch
37	7	0.9	211	1 IAP_ARATH	O64628 arabidopsis
38	7	0.9	211	1 YRAE_ECOLI	P45469 escherichia
39	7	0.9	213	1 PCPE_SYNEL	P50038 synecchococ
40	7	0.9	213	1 URE1_HELMO	P50044 helicobacte
41	7	0.9	238	1 PSB2_RAT	O63798 rattus norv
42	7	0.9	239	1 PSB2_MOUSE	P97372 mus musculu
43	7	0.9	271	1 EL2_RAT	P00774 rattus norv
44	7	0.9	280	1 PEG_OCBH	O8ep66 oceanobacti
45	7	0.9	284	1 LICH_VIBCH	O07350 vibrio chol
46	7	0.9	284	1 TPM1_BIOGL	P42636 biophalari
47	7	0.9	284	1 TPM2_BIOGL	P43689 biophalari
48	7	0.9	284	1 TPMW_TRICO	P15846 trichostion
49	7	0.9	293	1 LECT_CCATU	O39528 clatrastis
50	7	0.9	316	1 TAL_BUCAP	O8ka27 buchnera ap
51	7	0.9	323	1 PTPA_RABIT	O28717 oryctolagus
52	7	0.9	327	1 FIP1_YEAST	P45976 saccharomy
53	7	0.9	327	1 KAPR_DICDI	P05987 dictyostell
54	7	0.9	338	1 G3P2_AGAB1	P32636 agaricus bl
55	7	0.9	338	1 G3P_SCHMA	P20287 schistosoma
56	7	0.9	344	1 DCUP_AGRIS	O8ud16 agrobacteri
57	7	0.9	358	1 PTPA_HUMAN	O15257 homo sapien
58	7	0.9	386	1 PHEA_ECOLI	P07022 escherichia
59	7	0.9	387	1 YHFX_ECOLI	P45550 escherichia
60	7	0.9	394	1 PGK_THETN	O8r965 thermoaer
61	7	0.9	397	1 PGK_CHUTE	O8kael chlorobium
62	7	0.9	405	1 IF5_YEAST	P38431 saccharomy
63	7	0.9	415	1 RPA3_YEAST	O01080 saccharomy
64	7	0.9	416	1 HEM1_AERPE	O9y9j2 aeropyrum p
65	7	0.9	424	1 Y12A_ECOLI	P51026 escherichia
66	7	0.9	430	1 TRPB_HALNI	O9bco halobacteri
67	7	0.9	432	1 GSA_AERPE	O9y9i9 aeropyrum p
68	7	0.9	446	1 TBG_SCHUP	O9y882 schizosacch
69	7	0.9	446	1 TBG_SCHPO	P25295 schizosacch
70	7	0.9	450	1 V50K_BYDVP	P09516 barley yell
71	7	0.9	451	1 TBG1_HUMAN	P23358 homo sapien
72	7	0.9	451	1 TBG1_MOUSE	O92310 mus musculu
73	7	0.9	451	1 TBG2_HUMAN	O9nht3 homo sapien
74	7	0.9	451	1 TBG2_MOUSE	O8vck3 mus musculu
75	7	0.9	451	1 TBG2_XENLA	P23330 xenopus lae
76	7	0.9	455	1 YXX5_CAEEL	O18179 caenorhabdi
77	7	0.9	460	1 SELA_PASWJ	O9c66 pasteurrella
78	7	0.9	461	1 TBG1_EUPOC	P34786 euploies oc
79	7	0.9	461	1 TBG2_EUPOC	P90548 euploies oc
80	7	0.9	461	1 TBG2_NEUCR	P53377 neurospora
81	7	0.9	461	1 TR1B_HUMAN	P20333 homo sapien
82	7	0.9	462	1 TBG1_EUPCR	P54403 euploies cr
83	7	0.9	467	1 V51K_BMYVF	O09514 beet wester
84	7	0.9	471	1 V51K_BMYVG	P09515 beet wester
85	7	0.9	476	1 YHGF_NEIGO	O51062 neisseria g
86	7	0.9	479	1 VGLC_HSV2G	P03173 herpes simp
87	7	0.9	480	1 VGLC_HSV2H	P06475 herpes simp
88	7	0.9	480	1 VGLC_HSV2H	O89730 herpes simp
89	7	0.9	488	1 SUOX_HUMAN	P51887 homo sapien
90	7	0.9	491	1 K2M2_SHEEP	P15241 ovis aries
91	7	0.9	502	1 K2M3_SHEEP	P25691 ovis aries
92	7	0.9	508	1 V56K_PIRV1	P17525 potato leaf
93	7	0.9	508	1 V56K_PIRVW	P11126 potato leaf
94	7	0.9	509	1 GUAA_CLOPE	O8x146 clostridium
95	7	0.9	510	1 GUAA_CLOAB	O97748 clostridium
96	7	0.9	511	1 LYN_HUMAN	P07448 homo sapien
97	7	0.9	520	1 C84A_ARATH	O42600 arabidopsis
98	7	0.9	524	1 CP11_CANFA	P56590 canis fami
99	7	0.9	534	1 GCR2_YEAST	O01122 saccharomy
100	7	0.9	538	1 GLE1_YEAST	O12153 saccharomy
101	7	0.9	542	1 GUAA_SYNY3	P49557 synecchocyt
102	7	0.9	543	1 ZN08_HUMAN	P17098 homo sapien
103	7	0.9	570	1 PTI_BACSU	P08838 bacillus su
104	7	0.9	570	1 SYE_PYRAE	O8zn33 pyrobaculum
105	7	0.9	580	1 GGT_ECOLI	P18956 escherichia
106	7	0.9	588	1 TAN_ASFOR	P78581 aspergillus

107	7	0.9	589	1	MAYS YEAST	P40439	saccharomyc
108	7	0.9	589	1	SYFB HUMAN	O9nsd9	homo sapien
109	7	0.9	599	1	DNAX_MYCPU	O9nsy7	mycoplasma
110	7	0.9	613	1	GLMS_CHUTE	O8K938	c glucosami
111	7	0.9	615	1	DXS_CAVUE	O9p1h8	campylobact
112	7	0.9	627	1	DNAX_RICCN	O9z2j36	ricketsia
113	7	0.9	627	1	DNAX_RICPR	O9zdx9	ricketsia
114	7	0.9	649	1	PLBI_TORDE	O1121	torulaspore
115	7	0.9	651	1	PIAI_HUMAN	O75925	homo sapien
116	7	0.9	651	1	PIAI_MOUSE	O88907	mus musculu
117	7	0.9	657	1	PKN2_CLOAB	O971c2	clostridium
118	7	0.9	659	1	SPA2_SYNY3	P72587	synchocyst
119	7	0.9	677	1	SYM_PSEAB	O9hyct7	pseudomonas
120	7	0.9	688	1	DNAX_ANASP	O9z2j6	anaeana sp
121	7	0.9	688	1	DNAX_ANAVA	O05714	anaeana va
122	7	0.9	694	1	CNG3_HUMAN	O16281	homo sapien
123	7	0.9	725	1	YU76 YEAST	P47144	saccharomyc
124	7	0.9	738	1	VU47_HSV62	P52549	human herpe
125	7	0.9	740	1	YNO4 YEAST	P53907	saccharomyc
126	7	0.9	754	1	KB15_HUMAN	O9unp7	homo sapien
127	7	0.9	757	1	HUNB_DROSE	O62538	drosoophila
128	7	0.9	757	1	YHGF_NEIMA	P57072	neisseria m
129	7	0.9	757	1	YHGF_NEIMB	O51152	neisseria m
130	7	0.9	758	1	HUNB_DROME	P05084	drosoophila
131	7	0.9	759	1	HUNB_DROYA	O62541	drosoophila
132	7	0.9	764	1	YIS7 YEAST	P40566	saccharomyc
133	7	0.9	767	1	HUNB_DROOR	O62537	drosoophila
134	7	0.9	773	1	YHGF_ECOLI	P46837	escherichia
135	7	0.9	803	1	VE94_NPVAC	P08161	autographa
136	7	0.9	816	1	HUNB_DROVI	P13161	drosoophila
137	7	0.9	836	1	GYRA_MYCCE	P47250	mycoplasma
138	7	0.9	840	1	PHL1_HUMAN	P80108	homo sapien
139	7	0.9	842	1	EP2_SCHPO	O14660	schizosacch
140	7	0.9	860	1	GYRA_SYNY3	O55738	synchocyst
141	7	0.9	880	1	ENV_STVLM	P11267	simian immu
142	7	0.9	881	1	ENV_STVLM	P05884	simian immu
143	7	0.9	882	1	ENV_STVLM	P05885	simian immu
144	7	0.9	905	1	GYRA_RICCN	O92126	ricketsia
145	7	0.9	905	1	GYRA_RICPR	P41080	ricketsia
146	7	0.9	925	1	NI07_HUMAN	P57740	homo sapien
147	7	0.9	926	1	NI07_RAT	P52590	rattus norv
148	7	0.9	937	1	NU98_HUMAN	P52458	homo sapien
149	7	0.9	937	1	NU98_RAT	P49793	rattus norv
150	7	0.9	1193	1	LMG2_HUMAN	O13753	homo sapien
151	7	0.9	1203	1	SMC_WYCLE	O9c8t5	mycobacteri
152	7	0.9	1228	1	ALM5_ARATH	O9B993	arabidopsis
153	7	0.9	1314	1	SW11 YEAST	P09547	saccharomyc
154	7	0.9	1325	1	YAB6_SCHPO	O09847	schizosacch
155	7	0.9	1376	1	RPOB_RICTY	P77941	ricketsia
156	7	0.9	2144	1	GLT1 YEAST	O12680	saccharomyc
157	7	0.9	2175	1	HMCU_DROME	P10180	drosoophila
158	7	0.9	2547	1	FAFX_HUMAN	O93008	h probabile
159	7	0.9	2559	1	FAFX_MOUSE	P70398	m probable
160	7	0.9	3519	1	OL56 STRAT	O07017	streptomyc
161	7	0.9	3591	1	FHAB_BORPE	P12255	botdella
162	7	0.9	3828	1	TRX_DROVI	O24742	drosoophila
163	7	0.9	16	1	R16_VIBPR	O56715	vibriu proc
164	7	0.9	44	1	PSBF_SYNP7	O8K9P2	synchococc
165	7	0.9	50	1	RLJ3_FUSNN	O8rth9	tubobacteri
166	7	0.9	64	1	RKJ3_MARPO	P06392	marchantia
167	7	0.9	67	1	RLJ3_AERPE	O9yfer9	aeropyrum p
168	7	0.9	71	1	YVEA_VACCC	P21085	vaccinia vi
169	7	0.9	72	1	ZNPF_LYCVT	P19925	lymphocytic
170	7	0.9	76	1	ANRX_ANASP	O44141	anaeana sp
171	7	0.9	77	1	Y812_RICPR	O9zced4	ricketsia
172	7	0.9	80	1	CXO1_CONPL	O9zy20	conus pulic
173	7	0.9	81	1	RS16_CLOAB	O97197	clostridium
174	7	0.9	82	1	MLC1_DROMA	O24399	drosoophila
175	7	0.9	82	1	MLC1_DROSE	O24656	drosoophila
176	7	0.9	82	1	MLC1_DROTE	O24695	drosoophila
177	7	0.9	82	1	MLC1_DROVA	O24666	drosoophila
178	7	0.9	82	1	YFPD_METFE	P55509	methanother
179	7	0.9	83	1	GVPP_HALME	O02232	halobacteri

180	6	0.8	85	1	HKL6_MAIZE	P56664	zea mays (m
181	6	0.8	85	1	HKL7_MAIZE	P56665	zea mays (m
182	6	0.8	86	1	MLC1_DROSO	O44107	drosoophila
183	6	0.8	89	1	EAT1_ENTFA	O86101	enterococcu
184	6	0.8	89	1	EAT2_ENTFA	O96119	enterococcu
185	6	0.8	89	1	EATX_STREY	O57231	streptococc
186	6	0.8	90	1	RS16_LACIA	O9cfb2	lactococcus
187	6	0.8	90	1	WMJC_BACSU	P45863	bacillus su
188	6	0.8	93	1	CENA_BOVIN	P49449	bos taurus
189	6	0.8	93	1	L10K_RAT	O05310	rattus norv
190	6	0.8	94	1	DEF1_RAT	O62716	rattus norv
191	6	0.8	94	1	DEF2_RAT	O62717	rattus norv
192	6	0.8	95	1	DAF1_TRYCR	O26327	trypanosoma
193	6	0.8	95	1	GRX_RICPR	O9zdfb2	ricketsia
194	6	0.8	97	1	RL21_PYRAB	O9uzp1	pyrococcus
195	6	0.8	97	1	RL21_PYRHO	O74001	pyrococcus
196	6	0.8	100	1	NUOK_ECOLI	P33606	escherichia
197	6	0.8	100	1	TH10_MYCGA	O9f6p9	mycoplasma
198	6	0.8	101	1	NUIC_MAIZE	P11646	zea mays (m
199	6	0.8	101	1	NUIC_NEPOL	O9fkw2	nephrolepni
200	6	0.8	101	1	NUIC_ORYSA	P12128	oryza sativ

ALIGNMENTS

RESULT 1

7D3_DROME	STANDARD; PRT; 921 AA.
ID T2D3_DROME	P47825; P49845; Q879E0; Q9VUV7;
AC P47825;	Q879E0; Q9VUV7;
DT 01-FEB-1996 (Rel. 33, Created)	
DT 01-FEB-1996 (Rel. 33, Last sequence update)	
DT 15-SEP-2003 (Rel. 42, Last annotation update)	
DE Transcription initiation factor TFIID 110 kDa subunit (P110)	
DE (TAF11-110) (110 kDa TBP-associated factor).	
GN TAF4 OR TAF110 OR CG5444.	
OS Drosophila melanogaster (Fruit fly).	
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;	
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
OC Ephydroidea; Drosophilidae; Drosophila.	
OX NCBI_TaxID=7227;	
NP [1]	
NP SEQUENCE FROM N.A. (ISOFORM LONG), AND SEQUENCE OF 519-540; 597-616	
RP AND 857-874.	
RC TISSUE=Embryo;	
RX MEDLINE=93145326; PubMed=7678780;	
RA Hoey T., Weinzierl R.O.J., Gill G., Chen J.-L., Dynlacht B.D.,	
RA Tjian R.;	
RT "Molecular cloning and functional analysis of Drosophila TAF110	
RT reveal properties expected of coactivators.";	
RL Cell 72:247-260(1993).	
RM [2]	
RM SEQUENCE FROM N.A. (ISOFORM LONG), AND SEQUENCE OF 398-406; 520-540	
RP AND 860-877.	
RC TISSUE=Embryo;	
RX MEDLINE=93317591; PubMed=8327460;	
RA Kokubo T., Gong D.-W., Roeder R.G., Horikoshi M., Nakatani Y.;	
RT "The Drosophila 110-Kda transcription factor TFIID subunit directly	
RT interacts with the N-terminal region of the 230-Kda subunit.";	
RL Proc. Natl. Acad. Sci. U.S.A. 90:5896-5900(1993).	
RM [3]	
RM SEQUENCE FROM N.A.	
RP STRAIN=Berkley;	
RC MEDLINE=20196006; PubMed=10731132;	
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,	
RA Amaniatis P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,	
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,	
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,	
RA Brandon R.C., Rogers Y.-H.C., Blazey R.C., Champe M., Pfeiffer B.D.,	
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,	
RA Abfrit J.F., Agbayani A., An H.-J., Andrews-Piankovich C., Baldwin D.,	
RA Balles R.M., Basu A., Baxendale J., Bayraktarglu L., Beasley E.M.,	
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,	

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Cencer A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matcel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
 RA Palazzolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wattaman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195 (2000).
 [4]
 RN SEQUENCE FROM N.A. (ISOFORM SHORT).
 RP STRAIN=Berkley; TISSUE=embryo;
 RX MEDLINE=22426066; PubMed=12537569;
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
 RA George R.A., Guatin H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,
 RA Rubin G.M., Celinker S.E.;
 RT "A *Drosophila* full-length cDNA resource.";
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.6 (2002).
 CC -!- FUNCTION: TEIID IS A MULTIMERIC PROTEIN COMPLEX THAT PLAYS A
 CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS ACTIVATORS
 AND REPRESSORS. MAY FUNCTION AS A COACTIVATOR BY SERVING AS A SITE
 OF PROTEIN-PROTEIN CONTACT BETWEEN ACTIVATORS LIKE SPI (OR BTB)
 AND TEIID COMPLEX.
 CC -!- SUBUNIT: TEIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
 NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=Long;
 CC IsoId=P47825-1; Sequence=Displayed;
 CC Name=Short;
 CC IsoId=P47825-2; Sequence=VSP_004441;
 CC Note=No experimental confirmation available;
 CC -!- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
 CC
 CC -!- This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC
 CC EMBL, L06861; -, NOT ANNOTATED_CDS.
 DR EMBL, S63550; AAB27433.1; -
 DR EMBL, AE003528; AAF49536.1; -
 DR EMBL, AY069807; AAL39952.1; -
 DR PIR, A48184; A48184;
 DR TRANSFAC, T02121, -
 DR Flybase, FBgn010280; Tafa4.
 DR GO, GO:0005669; C:transcription factor TEIID complex; IPI.
 DR GO, GO:001251; I:general RNA polymerase II transcription fac. .; IPI.
 DR GO, GO:0006355; P:regulation of transcription, DNA-dependent; IPI.

DR GO, GO:0006367; P:transcription initiation from Pol II promoter; IPI.
 DR InterPro, IPR003894; TAF hom.
 DR SMART, SM00549; TAFH; 1.
 KW Transcription regulation; Nuclear protein; Alternative splicing.
 FT DOMAIN 66 82 POLY-GLN.
 FT DOMAIN 108 111 POLY-GLN.
 FT DOMAIN 259 265 POLY-GLN.
 FT VARSPLIC 138 207 Missing (in isoform Short).
 FT FTID=VSP_004441.
 FT V -> M (IN REF. 4).
 SQ CONFLICT 114 114
 SQ SEQUENCE 921 AA; 99338 MW; 27B6852859872767 CRC64;
 Query Match 1.1%; Score 9; DB 1; Length 921;
 Best Local Similarity 100.0%; Pred. No. 1.7;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 249 TINTSNNSNT 257
 DB 178 TINTSNNSNT 186
 RESULT 2
 Y029_NPVAC STANDARD; PRT; 71 AA.
 AC P41433;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Hypothetical 8.6 kDa protein in IAP1-SOD intergenic region.
 OS Autographa californica nuclear polyhedrosis virus (ACNPV).
 OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
 OC Nucleopolyhedrovirus.
 OK NCBI_TaxId=46015;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP STRAIN=C6;
 RX MEDLINE=94303173; PubMed=8030224;
 RA Ayres M.D., Howard S.C., Kuzio J., Lopez-Ferber M., Possee R.D.;
 RT "The complete DNA sequence of Autographa californica nuclear
 polyhedrosis virus.";
 RL Virology 202:586-605 (1994).
 CC -!- SIMILARITY: TO CORRESPONDING ORF IN OPNPV.
 CC
 CC -!- This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC
 CC EMBL, L22858; AAA66659.1; -
 DR EMBL, E72853; E72853.
 KW Hypothetical protein.
 SQ SEQUENCE 71 AA; 8569 MW; ADF85AC68E16DD3 CRC64;
 Query Match 1.0%; Score 8; DB 1; Length 71;
 Best Local Similarity 100.0%; Pred. No. 1.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 130 EIRNRKQ 137
 DB 20 EIRNRKQ 27
 RESULT 3
 SC17_NEUCR STANDARD; PRT; 292 AA.
 AC Q9P6A5;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Probable vesicular-fusion protein sc17 homolog.

```

GN B1D1.150.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=74-OR23-1A / FGSC 987;
RX PubMed=12655011;
RA Mannhaupt G., Montlone C., Haase D., Mewes H.-W., Aign V.,
RA Honeisel J.D., Fartmann B., Nyakatura G., Kempken F., Walter J.,
RA Schulte U.;
RT "What's in the genome of a filamentous fungus? Analysis of the
RT Neurospora genome sequence."
RL Nucleic Acids Res. 31:1944-1954(2003).
CC -1- FUNCTION: REQUIRED FOR VESICULAR TRANSPORT BETWEEN THE ENDOPLASMIC
CC RETICULUM AND THE GOLGI APPARATUS (By similarity).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC PERIPHERAL MEMBRANE PROTEIN (By
CC similarity).
CC -1- SIMILARITY: BELONGS TO THE SNAP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL355927; CAB91264.1; -.
DR PIR; T49361; T49361.
DR InterPro; IPR000744; NSF_attach.
DR Pfam; PF02071; NSF; 2
DR PRINTS; PR00448; NSFATTACHMNT.
DR Transport; Protein transport; Endoplasmic reticulum; Golgi stack.
SQ SEQUENCE 292 AA; 32805 MW; C9BD75D8128E19E7 CRC64;

Query Match 1.0%; Score 8; DB 1; Length 292;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 NDGAVALA 157
Db 150 NDGAVALA 157

RESULT 4
AAC2_DICD1 STANDARD; PRT; 448 AA.
ID AAC2_DICD1
AC P14196;
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE AAC-rich mRNA clone AAC11 protein (Fragment).
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90066348; PubMed=2511421;
RA Shaw D.R., Richter H., Giorda R., Ohmachi T., Emis H.L.;
RT "Nucleotide sequences of Dictyostelium discoideum developmentally
RT regulated cDNAs rich in (AAC) imply proteins that contain clusters of
RT asparagine, glutamine, or threonine."
RL MoJ. Genet. 218:453-459(1989).
CC -1- DEVELOPMENTAL STAGE: THE CONCENTRATION OF AAC-RICH MNAS IS LOW
CC IN DORMANT SPORES AND GROWING CELLS, BUT INCREASES DURING
CC SPORE-GERMINATION AND MULTICELLULAR DEVELOPMENT.
CC -1- MISCELLANEOUS: SEVERAL PROTEINS DERIVE FROM AAC-RICH MRNA, WHICH,
CC DUE TO A FRAMESHIFT ALSO HAVE ACA AND CAA CODONS AND THUS ARE
CC ASN-, THR- OR GIN-RICH.
CC -1- SIMILARITY: Contains 4 A.T hook DNA-binding repeats.
CC -----

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X16522; CAA34529.1; -.
DR PIR; S05355; S05355.
DR Dictydb; DD05006; -.
DR InterPro; IPR000637; AT_hook.
DR InterPro; IPR000116; Highmobility_1Y.
DR Pfam; PF02178; AT_hook; 4.
DR PRINTS; PR00929; ATHOOK.
DR PRODOM; PD005593; Highmobility_1Y; 1.
DR SMART; SM00384; AT_hook; 4.
KW DNA-binding; Repeat.
FT NON TER 1
FT DOMAIN 5 25 GLN-RICH.
FT DOMAIN 115 144 ASN-RICH.
FT DOMAIN 268 384 ASN-RICH.
SQ SEQUENCE 448 AA; 48636 MW; D8FD80D910D99817 CRC64;

Query Match 1.0%; Score 8; DB 1; Length 448;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 250 NTSNNSNT 257
Db 378 NTSNNSNT 385

RESULT 5
IRF7_HUMAN STANDARD; PRT; 503 AA.
ID IRF7_HUMAN
AC Q92985; O00331; O00332; O00333; O75924;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Interferon regulatory factor 7 (IRF-7).
GN IRF7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM A).
RC TISSUE=Spleen;
RA Grossman A., Nicholl J., Antonio L., Luethy R., Suggs S.,
RA Sutherland G.R., Mak T.W.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS A; B AND C).
RX MEDLINE=97459673; PubMed=9315633;
RA Zhang L., Pagano J.S.;
RT "IRF-7, a new interferon regulatory factor associated with Epstein-
RT Barr virus latency."
RL MoJ. Cell. Biol. 17:5748-5757(1997).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM D).
RX MEDLINE=99003279; PubMed=9786932;
RA Au W.-C., Moore P.A., Lafleur D.W., Tombal B., Pitla P.M.;
RT "Characterization of the interferon regulatory factor-7 and its
RT potential role in the transcription activation of interferon A
RT genes."
RL J. Biol. Chem. 273:29210-29217(1998).
CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR. BINDS TO THE INTERFERON-
CC STIMULATED RESPONSE ELEMENT (ISRE) IN IFN PROMOTERS AND IN THE Q
CC PROMOTER (QP) OF EBV NUCLEAR ANTIGEN-1 (EBNA1).
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;

```

```

CC      Name=A;
CC      IsoId=Q92985-1; Sequence=Displayed;
CC      Name=B; Synonyms=Beta;
CC      IsoId=Q92985-2; Sequence=VSP_002760;
CC      Name=C; Synonyms=Gamma;
CC      IsoId=Q92985-3; Sequence=VSP_002758, VSP_002759;
CC      Name=D; Synonyms=H;
CC      IsoId=Q92985-4; Sequence=VSP_002757;
CC      TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN SPLEEN, THYMUS, AND
CC      PERIPHERAL BLOOD LEUCOCYTES.
CC      -1- SIMILARITY: Belongs to the IRF family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; U73036; AAB17190.1; -
CC      EMBL; U53830; AAB80686.1; -
CC      EMBL; U53831; AAB80688.1; -
CC      EMBL; U53832; AAB80690.1; -
CC      EMBL; AF076494; AAC70999.1; -
CC      HSSP; P23906; 2IRF.
CC      TRANSFAC; T04674; -.
CC      Genew; HGNC:6122; IRF7.
CC      MIM; 605047; -.
CC      GO; GO:0005737; Cytoplasm; TAS.
CC      GO; GO:0005634; C:nucleus; TAS.
CC      GO; GO:0003704; F:specific RNA polymerase II transcription fa. .; TAS.
CC      GO; GO:0000122; P:negative regulation of transcription from P. .; TAS.
CC      GO; GO:0009615; P:response to viruses; TAS.
CC      InterPro; IPR001346; IRF.
CC      Pfam; PF00605; IRF.1.
CC      PRINTS; PR00267; INTERFERGCT.
CC      PRODOM; PD002355; IRF.1.
CC      SMART; SM00348; IRF.1.
CC      PROSITE; PS00601; IRF.1.
CC      Transcription regulation; DNA-binding; Nuclear protein; Activator;
CC      Alternative splicing.
CC      DNA BIND 13 122 TRYPTOPHAN PENTAD REPEAT.
CC      VARSPLIC 1 6 MALAPE -> MPVPERPAGDPSPRQTR (in isoform
CC      D).
CC      VARSPLIC 152 164 /FTid=VSP_002757.
CC      GDBPGLAHTHA -> AQSLLGSCCTGQ (in
CC      isoform C).
CC      VARSPLIC 165 503 /FTid=VSP_002758.
CC      Missing (in isoform C).
CC      VARSPLIC 228 256 /FTid=VSP_002759.
CC      Missing (in isoform B).
CC      CONFLICT 179 179 /FTid=VSP_002760.
CC      CONFLICT 412 412 E -> K (IN REF. 2).
CC      SEQUENCE 503 AA; 54278 MW; AAG39E0E272727C CRC64;

```

```

Query Match      1.0%; Score 8; DB 1; Length 503;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      357 PBPPOPA 364
      |||||
Db      244 PBPPOPA 251

```

```

RESULT 6
LON MYCPN      STANDARD;      PRT;      795 AA.
ID  LON MYCPN
AC  P78025;
DT  01-NOV-1997 (Rel. 35, Created)
DT  01-NOV-1997 (Rel. 35, Last sequence update)
DT  28-FEB-2003 (Rel. 41, Last annotation update)

```

```

DE      ATP-dependent protease Ia (EC 3.4.21.53).
LON OR MPN332 OR MP504.
OS      Mycoplasma pneumoniae.
OC      Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX      NCBI_Taxid=2104;
RN      (1)
RP      SEQUENCE FROM N.A.
RC      STRAIN=ATCC 29342 / M129;
RX      MEDLINE=97105885; PubMed=8948633;
RA      Himmelfrich R., Hilbert H., Piegens H., Pirkl E., Li B.-C.,
RA      Herrmann R.;
RT      "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT      pneumoniae."
RL      Nucleic Acids Res. 24:4420-4449 (1996).
CC      -1- FUNCTION: DEGRADATES SHORT-LIVED REGULATORY AND ABNORMAL PROTEINS
CC      IN PRESENCE OF ATP. DEGRADATES THE REGULATORY PROTEINS RCSA AND
CC      SULA. HYDROLYZES TWO ATPs FOR EACH PEPTIDE BOND CLEAVED IN THE
CC      PROTEIN SUBSTRATE (BY SIMILARITY).
CC      -1- CATALYTIC ACTIVITY: Hydrolysis of large proteins such as globin,
CC      casein and denatured serum albumin, in presence of ATP.
CC      -1- SUBUNIT: Homotetramer (By similarity).
CC      -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC      -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S16.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; AE000050; AAB96152.1; -.
CC      PIR; S73830; S73830.
CC      MEROPS; S16.004; -.
CC      InterPro; IPR003593; AAA ATPase.
CC      InterPro; IPR003959; AAA ATPase_cent.
CC      InterPro; IPR003111; LON.
CC      InterPro; IPR001984; Lon_endopep.
CC      InterPro; IPR004815; Lon_fam.
CC      Pfam; PF00004; AAA; 1.
CC      Pfam; PF02190; LON; 1.
CC      Pfam; PF03562; Lon_C.1.
CC      PRINTS; PR00830; ENDOLAPYASE.
CC      SMART; SM00382; AAA; 1.
CC      SMART; SM00464; LON; 1.
CC      TIGRPFAMS; TIGR00763; Lon; 1.
CC      PROSITE; PS01046; LON_SER; 1.
CC      Hydrolase; Serine protease; ATP-binding; Complete proteome.
CC      FT DOMAIN 304 308 POLY-SER.
CC      NP BIND 379 386 ATP (POTENTIAL).
CC      ACT_SITE 702 702 BY SIMILARITY.
CC      SEQUENCE 795 AA; 90203 MW; 7E7855082060C891 CRC64;

```

```

Query Match      1.0%; Score 8; DB 1; Length 795;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      466 ALDKLER 473
      |||||
Db      127 ALDKLER 134

```

```

RESULT 7
Y240 ARATH      STANDARD;      PRT;      861 AA.
ID  Y240 ARATH
AC  O22224; O9ACF2;
DT  16-OCT-2001 (Rel. 40, Created)
DT  28-FEB-2003 (Rel. 41, Last sequence update)
DT  28-FEB-2003 (Rel. 41, Last annotation update)
DE  Protein At2g41620.
GN  AT2G41620 OR T32C6.14.
OS  Arabidopsis thaliana (Mouse-ear cress).

```

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eucosids II; Brassicales; Brassicaceae; Arabidopsis.
 CX NCB1_TaxId=3702;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20083487; PubMed=10617197;
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
 RA Buell C.R., Ketchum K.A., Lee J.C., Renning C.M., Koo H.L.,
 RA Moffat K.S., Cronin L.A., Shen M., Pal G., Van Aken S., Umayam L.,
 RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
 RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
 RA Newman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
 RA Venter J.C.,
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
 RT thaliana.";
 RL Nature 402:761-768 (1999).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
 RT "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
 RT SSP consortium (Salk/Stanford/GEC).";
 RL Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.
 RN (3)
 RP IDENTIFICATION ON 2D-GELS.
 RC STRAIN=cv. Columbia;
 RA Sarazin B., Tonella L., Marques K., Paesano S., Chane-Favre L.,
 RA Heller M., Sanchez J.-C., Hochstrasser D.F., Thiellement H.;
 RL Unpublished observations (SEP-2000).
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AC002510; AAB84345.2; -;
 DR EMBL: AY034907; AAC59414.1; -;
 DR PFI: PF04097; NAK: 1.
 SQ SEQUENCE 861 AA; 96615 MW; 9F53F0BF013D673 CRC64;
 Query Match 1.0%; Score 8; DB 1; Length 861;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 41 REGINAEO 48
 Db 69 REGINAEO 76
 RESULT 8
 HDC DROME STANDARD; PRT; 1080 AA.
 ID HDC DROME
 AC Q5N2M8; Q24480; Q9VAB4;
 DT 16-OCT-2001 (Rel. 40; Created)
 DT 16-OCT-2001 (Rel. 40; Last sequence update)
 DT 28-FEB-2003 (Rel. 41; Last annotation update)
 DE Headcase protein [Contains: Headcase short protein].
 GN HDC OR CG15532.
 OS Drosophila melanogaster (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 CX NCB1_TaxId=7227;
 RN (1)
 RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
 RC TISSUE=Embryo;
 RX MEDLINE=96171720; PubMed=8575315;

RA Weaver T.A., White R.A.;
 RT "Headcase, an imaginal specific gene required for adult morphogenesis
 RT in Drosophila melanogaster.";
 RL Development 121:4149-4160 (1995).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Goeayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champs M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Mklos G.L.G.,
 RA Abril J.F., Agbayan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 RA Butris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasner K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.B., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclet J.M.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Part V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skipski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svrtskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye Y., Yeh R.-F., Zaveri J.S., Zhan W., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195 (2000).
 RN (3)
 RP PARTIAL SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
 RC TISSUE=Embryo;
 RX MEDLINE=98198453; PubMed=9531534;
 RA Steneberg P., Englund C., Kiroham J., Weaver T.A., Samkovits C.;
 RT "Translational readthrough in the hdc mRNA generates a novel branching
 RT inhibitor in the drosophila trachea.";
 RL Genes Dev. 12:956-967 (1998).
 CC CC
 CC -I- FUNCTION: REQUIRED FOR IMAGINAL CELL DIFFERENTIATION. MAY BE
 CC INVOLVED IN HORMONAL RESPONSIVENESS DURING METAMORPHOSIS. INVOLVED
 CC IN AN INHIBITORY SIGNALING MECHANISM TO DETERMINE THE NUMBER OF
 CC CELLS THAT WILL FORM UNICELLULAR SPROUTS IN THE TRACHEA. REGULATED
 CC BY TRANSCRIPTION FACTOR E5G. THE LONGER HDC PROTEIN IS COMPLETELY
 CC FUNCTIONAL AND THE SHORTER PROTEIN CARRIES SOME FUNCTION.
 CC CC
 CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
 CC CC
 CC -I- TISSUE SPECIFICITY: EXPRESSED IN A SUBSET OF TRACHEAL FUSION CELLS FROM
 CC STAGE 14 TO THE END OF EMBRYOGENESIS IN METAMERES 2-9, LATERAL
 CC TRUNK AND VENTRAL ANASTOMOSES.
 CC CC
 CC -I- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR UAA OCCURS BETWEEN
 CC CODONS FOR ALA-650 AND HIS-652. READTHROUGH IS NOT ALWAYS
 CC SUPPRESSED AS THE SHORTER PROTEIN IS MORE ABUNDANT.
 CC CC
 CC -I- CAUTION: Ref.2 sequence differs from that shown due to erroneous
 CC gene model prediction.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

RX	MEDLINE=98295987; PubMed=9634230;
RA	Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA	Gordon S.V., Eglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA	Baddock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA	Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA	Hornsey T., Jaegle K., Krogh A., McLean J., Moule S., Murphy L.,
RA	Oliver S., Osborne J., Quail M.A., Rajendram M.A., Rogers J.,
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA	Sullivan J.E., Taylor K., Whitehead S., Batrell B.G.;
RT	"Deciphering the Biology of Mycobacterium tuberculosis from the
RT	complete genome sequence." ;
RL	Nature 393:537-544(1998).
RN	[3]
RP	SEQUENCE FROM N.A.
RC	STRAIN=CDC 1551 / Oshkosh;
RA	Pletschmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA	Fischeron J., Deboy R., Dodson R., Gwin M.L., Hatt D., Hickey E.,
RA	Kolony J.F., Nelson W.C., Unayama L.A., Ermolaeva M.D., Salzberg S.L.,
RA	Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA	Bislat W.;
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and
RT	laboratory strains." ;
RL	Submitted (APR-2001) to the EMBL/GenBank/DBE databases.
CC	- FUNCTION: PLAYS AN IMPORTANT ROLE IN CHROMOSOME STRUCTURE AND
CC	PARTITIONING. ESSENTIAL FOR CHROMOSOME PARTITION (BY SIMILARITY).
CC	- SIMILARITY: BELONGS TO THE SMC FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration -
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; AI144609; CAC93884.1; --
DR	EMBL; Z74697; CA989882.1; ALT_INIT.
DR	EMBL; AE007121; AAK47317.1; --
DR	TIGR; MT2990; --
DR	Tuberculinist; RV2922C; --
DR	InterPro: IPR003439; ABC transporter.
DR	InterPro: IPR003405; SMC_C.
DR	InterPro: IPR003395; SMC_N.
DR	Pfam; PF02483; SMC_C; 1.
DR	Pfam; PF02463; SMC_N; 1.
KW	ATP-binding; Coiled coil; Complete proteome.
NP	BIND 31 ATP (POTENTIAL).
FT	DOMAIN 38
FT	DOMAIN 167 289 COILED COIL (POTENTIAL).
FT	DOMAIN 330 499 COILED COIL (POTENTIAL).
FT	DOMAIN 659 922 COILED COIL (POTENTIAL).
FT	DOMAIN 899 929 COILED COIL (POTENTIAL).
FT	DOMAIN 979 1038 COILED COIL (POTENTIAL).
SO	SEQUENCE 1205 AA; 130637 MW; ASB2A813BS58EACF3 CRC64;
OY	203 SEIAAEA 210
Db	672 SEIAAEA 679
Query Match	1.0%; Score 8; DB 1; Length 1205;
Best Local Similarity	100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches	0; Indels 0; Gaps 0;
RESULT 10	
EVPL_HUMAN	STANDARD; PRT; 2033 AA.
ID_EVPL_HUMAN	Q92817;
DT 16-OCT-2001	(Rel. 40, Created)
DT 16-OCT-2001	(Rel. 40, Last sequence update)
DT 28-FEB-2003	(Rel. 41, Last annotation update)
DE Envoplakin (210 kDa paranecoplastic pemphigus antigen) (p210) (210 kDa	
DE consitied envelope precursor).	
NN EVPL.	

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 CX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=keratinocytes;
 RX MEDLINE=96326676; PubMed=8707850;
 RA Ruhrberg C., Hajibagheri M.A.N., Simon M., Dooley T.P., Watt F.M.,
 RT "Envoplakin, a novel precursor of the cornified envelope that has
 RL homology to desmoplakin.";
 J. Cell Biol. 134:715-729(1996).
 (2)
 RN SEQUENCE FROM N.A.
 RP MEDLINE=99339988; PubMed=10409435;
 RA Risk J.M., Ruhrberg C., Hennes H.-C., Mills H.S., Di Colandrea T.,
 RA Evans K.E., Ellis A., Watt F.M., Bishop D.T., Spurr N.K.,
 RA Stevens H.P., Leigh I.M., Reis A., Kelsell D.P., Field J.K.,
 RT "Envoplakin, a possible candidate gene for focal NPPK/esophageal
 cancer (TOC): the integration of genetic and physical maps of the TOC
 region on 17q25.";
 Genomics 59:234-242(1999).
 RL Genomics 59:234-242(1999).
 CC -1- FUNCTION: COMPONENT OF THE CORNIFIED ENVELOPE OF KERATINOCYTES.
 CC MAY LINK THE CORNIFIED ENVELOPE TO DESMOSOMES AND INTERMEDIATE
 CC FILAMENTS.
 CC -1- SUBUNIT: MAY FORM A HOMODIMER OR A HETERODIMER WITH PPL.
 CC -1- SUBCELLULAR LOCATION: COLOCALIZED WITH DSP AT DESMOSOMES AND ALONG
 CC INTERMEDIATE FILAMENTS.
 CC -1- TISSUE SPECIFICITY: EXCLUSIVELY EXPRESSED IN STRATIFIED SQUAMOUS
 CC EPITHELIA.
 CC -1- INDUCTION: DURING DIFFERENTIATION OF EPIDERMAL KERATINOCYTES.
 CC -1- SIMILARITY: Contains 7 plectrin repeats.
 CC -1- SIMILARITY: Contains 1 spectrin repeat.
 CC -1- SIMILARITY: Belongs to the plakin or cytolinker family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U53786; AAC64662.1; -;
 DR EMBL; U72849; AAD00186.1; -;
 DR EMBL; U72843; AAD00186.1; JOINED.
 DR EMBL; U72845; AAD00186.1; JOINED.
 DR EMBL; U72846; AAD00186.1; JOINED.
 DR EMBL; U72847; AAD00186.1; JOINED.
 DR EMBL; U72848; AAD00186.1; JOINED.
 DR Genew; HGNC:3503; EVPL.
 DR MIM; 601590; -;
 DR GO; GO:0009506; C:plasmodesma; TAS.
 DR GO; GO:0005198; F:structural molecule activity; TAS.
 DR GO; GO:0008544; P:epidermal differentiation; TAS.
 DR InterPro; IPR001101; Plectrin_repeat.
 DR Pfam; PF00681; Plectrin_3.
 DR SMART; SM00250; Plectr; 8.
 DR Keratinization; Repeat; Coiled coil; Cytoskeleton; Structural protein.
 FT DOMAIN 1 841 GLOBIULAR 1.
 FT DOMAIN 2 1673 CENTRAL FIBROUS ROD DOMAIN.
 FT DOMAIN 3 1674 GLOBIULAR 2.
 FT DOMAIN 4 1675 4 X 4 AA TANDEM REPEATS OF K-G-S-P.
 FT REPEAT 229 330 SPECTRIN.
 FT REPEAT 845 1135 COILED COIL (POTENTIAL).
 FT REPEAT 1185 1226 PLECTRIN 1.
 FT REPEAT 1678 1713 PLECTRIN 2.
 FT REPEAT 1818 1855 PLECTRIN 3.
 FT REPEAT 1856 1893 PLECTRIN 4.
 FT REPEAT 1894 1931 PLECTRIN 5.
 FT REPEAT 1932 1969 PLECTRIN 6.
 FT REPEAT 1970 2007 PLECTRIN 7.
 FT SEQUENCE 2033 AA; 231616 MW; B8DC6E2B52221938 CRC64;

Query Match 1.0%; Score 8; DB 1; Length 2033;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 731 LKAEVLL 738
 DB 87 LKAEVLL 94
 RESULT 11
 EVPL_MOUSE STANDARD: PRT, 2035 AA.
 ID EVPL_MOUSE
 AC 09D952;
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Envoplakin (p210) (210 kDa cornified envelope precursor).
 GN EVPL.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CX NCBI_TaxID=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/SV;
 RX MEDLINE=20347896; PubMed=10747979;
 RA Maatta A., Ruhrberg C., Watt F.M.,
 RT "Structure and regulation of the envoplakin gene.";
 RL J. Biol. Chem. 275:19857-19865(2000).
 (2)
 RP SEQUENCE OF 1860-2035 FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Pancreas;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Aizawa T., Hara A., Fukunishi Y., Konno H., Aachi J., Fukuda S.,
 RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nakado I., Pesole G., Quackenbush J.,
 RA Schmitt L.M., Steubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,
 RA Lyons P., Marchionni L., Washima J., Wazarelli U., Wombers P.,
 RA Nordone P., Ring B., Schoenbach C., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:665-690(2001).
 CC -1- FUNCTION: COMPONENT OF THE CORNIFIED ENVELOPE OF KERATINOCYTES.
 CC MAY LINK THE CORNIFIED ENVELOPE TO DESMOSOMES AND INTERMEDIATE
 CC FILAMENTS.
 CC -1- SUBUNIT: MAY FORM A HOMODIMER OR A HETERODIMER WITH PPL.
 CC -1- SUBCELLULAR LOCATION: COLOCALIZED WITH DSP AT DESMOSOMES AND ALONG
 CC INTERMEDIATE FILAMENTS (BY SIMILARITY).
 CC -1- SIMILARITY: Contains 7 plectrin repeats.
 CC -1- SIMILARITY: Contains 1 spectrin repeat.
 CC -1- SIMILARITY: Belongs to the plakin or cytolinker family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AJ309317; CAC38864.2; -;

DR EMBL; AJ319607; CAC38864.2; JOINED.
 DR EMBL; AJ319608; CAC38864.2; JOINED.
 DR EMBL; AJ319609; CAC38864.2; JOINED.
 DR EMBL; AJ319610; CAC38864.2; JOINED.
 DR EMBL; AJ319611; CAC38864.2; JOINED.
 DR EMBL; AJ319612; CAC38864.2; JOINED.
 DR EMBL; AJ319613; CAC38864.2; JOINED.
 DR EMBL; AK007353; -; NOT_ANNOTATED_CDS.
 DR MGD; MGI:107507; Eyp1.
 DR InterPro; IPR001101; Plectin_repeat.
 DR Pfam; PF00681; plectin; 3.
 DR Pfam; PF00435; spectrin; 2.
 DR SMART; SM00250; PLEC; 8.
 DR Keratinization; Repeat; Coiled coil; Cytoskeleton; Structural protein.
 FT DOMAIN 1 841
 FT DOMAIN 1675 2035
 FT DOMAIN 12 28
 FT REPEAT 229 330
 FT REPEAT 842 1664
 FT REPEAT 1186 1227
 FT REPEAT 1679 1714
 FT REPEAT 1819 1856
 FT REPEAT 1857 1894
 FT REPEAT 1895 1932
 FT REPEAT 1933 1970
 FT REPEAT 1971 2008
 FT REPEAT 1861 1864
 FT CONFLICT 2035 AA; 23217 MW; EB4A1D9CAED9641F CRC64;
 SQ SEQUENCE

Query Match
 Best Local Similarity 1.0%; Score 8; DB 1; Length 2035;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 731 LKEAEVLL 738
 Db 87 LKEAEVLL 94

RESULT 12
 Y872_PYRAB STANDARD; PRT; 70 AA.
 AC Q9VOB8;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Hypothetical protein PYRAB08720.
 GN PYRAB08720 OR PAB8218.
 OS Pyrococcus abyssi.
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
 CC Pyrococcus;
 CC NCBI_TaxID=29292;
 RN NCB1
 RP SEQUENCE FROM N.A.
 RC STRAIN=GES / Orsay;
 RX PubMed=12622808;
 RA Cohen G.N., Barbe V., Flament D., Galperin M., Helling R., Lecompte O.,
 Poch O., Frieur D., Querellou J., Ripp R., Thierry J.-C.,
 Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.;
 RA "An integrated analysis of the genome of the hyperthermophilic
 RT archaeon Pyrococcus abyssi";
 RL Mol. Microbiol. 47:1495-1512(2003).
 CC - SIMILARITY: BELONGS TO THE UPF0165 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AJ248285; CAB49786.1; -

DR PIR; A75134; A75134.
 DR InterPro; IPR002807; DUF104.
 DR Pfam; PF01954; DUF104; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 70 AA; 8009 MW; 4BDE6021B7D8032 CRC64;

Query Match
 Best Local Similarity 0.9%; Score 7; DB 1; Length 70;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 478 STNKEKL 484
 Db 34 STNKEKL 40

RESULT 13
 CH10_MYCBO STANDARD; PRT; 99 AA.
 ID CH10_MYCBO
 AC P15020;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 10 kDa chaperonin (Protein Cpn10) (groES protein) (Immunogenic protein
 DE MPB57).
 DE GROS OR GROES OR MOBP.
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 CC NCBI_TaxID=1765;
 RN NCB1
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-20.
 RC STRAIN=BCG;
 RX MEDLINE=89052868; PubMed=3056744;
 RA Yamaguchi R., Matsuo K., Yamazaki A., Nagai S., Terasaka K.,
 Yamada T.;
 RT "Immunogenic protein MPB57 from Mycobacterium bovis BCG: molecular
 RT cloning, nucleotide sequence and expression.";
 RL FEBS Lett. 240:115-117(1988).
 CC - FUNCTION: Binds to Cpn60 in the presence of Mg-ATP and suppresses
 CC the ATPase activity of the latter.
 CC - SUBUNIT: Heptamer of 7 subunits arranged in a ring (By
 CC similarity).
 CC - SIMILARITY: BELONGS TO THE GROES CHAPERONIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X13970; CAA32149.1; -
 CC EMBL; M35389; AAA25365.1; -
 CC PIR; S01784; BWMY7B.
 DR HSRP; P05380; IAON.
 DR HAMAP; MF_00580; -; 1.
 DR InterPro; IPR001476; Chaperin_Cpn10.
 DR Pfam; PF00166; Cpn10; 1.
 DR PRINTS; PR00297; CHAPERONIN10.
 DR PRODOM; PD000566; Chaperin Cpn10; 1.
 DR PROSITE; PS00681; CHAPERONIN_CPN10; 1.
 KW Chaperone; Antigen; Heat shock.
 FT INIT MER
 FT SEQUENCE 99 AA; 10819 MW; 1DCD5E2199447AF7 CRC64;

Query Match
 Best Local Similarity 0.9%; Score 7; DB 1; Length 99;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 605 GEKRIPL 611
 |||||

DB 54 GEXRPL 60

RESULT 14

ID CH10_MYCTU STANDARD; PRT; 99 AA.

AC P09621;

DT 01-MAR-1989 (Rel. 10, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE 10 kDa chaperonin (Protein Cpn10) (groES protein) (BCG-A heat shock protein) (10 kDa antigen).

GN GROES OR GROES OR MOBP OR CPN10 OR RV3418C OR MT3527 OR MYCTU8.11.

OS Mycobacterium tuberculosis.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.

OC NCBI_TaxID=1773;

OX [1]

RP SEQUENCE FROM N.A.

RC STRAIN=H37RV;

RX MEDLINE=89016584; PubMed=2302558;

RA Baird P.N., Hall L.M., Coates A.R.M.;

RT "A major antigen from Mycobacterium tuberculosis which is homologous to the heat shock proteins groES from E. coli and the hspA gene product of *Coxiella burnetii*."

RL Nucleic Acids Res. 16:9047-9047 (1988).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=H37RV;

RX MEDLINE=90095443; PubMed=2480990;

RA Baird P.N., Hall L.M.C., Coates A.R.M.;

RT "Cloning and sequence analysis of the 10 kDa antigen gene of Mycobacterium tuberculosis."

RL J. Gen. Microbiol. 135:931-939 (1989).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=Erdmann;

RX MEDLINE=89160258; PubMed=2564178;

RA Shinnick T.M., Plikaytis B.P., Hyche A.D., van Lingham R.M., Walker L.L.;

RT "The Mycobacterium tuberculosis BCG-a protein has homology with the *Escherichia coli* GroES protein."

RL Nucleic Acids Res. 17:1254-1254 (1989).

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN=Erdmann;

RX MEDLINE=92129332; PubMed=7681982;

RA Kong T.H., Coates A.R.M., Butcher P.D., Hickman C.J., Shinnick T.M.;

RT "Mycobacterium tuberculosis expresses two chaperonin-60 homologs."

RL Proc. Natl. Acad. Sci. U.S.A. 90:2608-2612 (1993).

RN [5]

RP SEQUENCE FROM N.A.

RC STRAIN=H37RV;

RX MEDLINE=98295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekela F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;

RT "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence."

RL Nature 393:537-544 (1998).

RN [6]

RP SEQUENCE FROM N.A.

RC STRAIN=CDC 1551 / Oshkosh;

RX STRAIN=CDC 1551 / Oshkosh;

RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O., Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E., Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L., Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A., Bishai W.;

RT "Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains."

RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

RN [7]

RP SEQUENCE OF 1-15.

RX MEDLINE=92176646; PubMed=1371791;

RA Barnes P.F., Mehra V., Riviere B., Fong S.J., Brennan P.J.;

RA Voegtlin M.S., Minden P., Houghten R.A., Bloom B.R., Modlin R.L.;

RT "Immunoreactivity of a 10-kDa antigen of Mycobacterium tuberculosis."

RL J. Immunol. 148:1835-1840 (1992).

CC -1- FUNCTION: Binds to Cpn60 in the presence of Mg-ATP and suppresses the ATPase activity of the latter.

CC -1- SUBUNIT: Heptamer of 7 subunits arranged in a ring (By similarity).

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

CC -1- SIMILARITY: BELONGS TO THE GROES CHAPERONIN FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X60350; CAA42908.1; -

DR EMBL; M25258; AAA25340.1; -

DR EMBL; X13739; CAA32003.1; -

DR EMBL; Z77165; CAB01005.1; -

DR EMBL; AE007158; AAK47865.1; -

DR PIR; S01381; BWMYBA.

DR PDB; 1HX5; 08-AUG-01.

DR PDB; 1JH2; 05-SEP-01.

DR TIGR; MT3527; -

DR Tuberculaet; RV3418C; -

DR HAMAP; MF_00580; -; 1.

DR InterPro; IPR001476; Chaperonin_Cpn10.

DR Pfam; PF00166; Cpn10; 1.

DR PRINTS; PR00297; CHAPERONIN10.

DR ProDom; PD000566; Chaperonin_Cpn10; 1.

DR PROSITE; PS00681; CHAPERONINS_CPN10; 1.

KW Chaperone; Antigen; Heat shock; Complete proteome; 3D-structure.

FT INIT MET 0

FT INT MET 0

FT SEQUENCE 99 AA; 10673 MW; 1DD128E75CF19AF7 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 99;

Best local Similarity 100.0%; Pred. No. 26;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 605 GEXRPL 611

DB 54 GEXRPL 60

RESULT 15

ID GLR1_SCHPO STANDARD; PRT; 101 AA.

AC O36032; Q9US58;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Glutaredoxin 1.

GN GRX1 OR SPACA10.20

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomycetes.

OC NCBI_TaxID=4896;

OX [1]

RP SEQUENCE FROM N.A.

RA Kawamukai M.;

RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.
 RA Kim H.-G., Cho Y.-W., Park E.-H., Lim C.-J.
 RT "Characterization of cDNA encoding thioltransferase (glutaredoxin)
 RL from Schizosaccharomyces pombe." ;
 RN Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 [3]
 RP SEQUENCE FROM N.A.
 RA Cho Y.-W., Kim H.-G., Lim C.-J.
 RT "Isolation and expression of the genomic DNA encoding thioltransferase
 RL (glutaredoxin) from Schizosaccharomyces pombe." ;
 RN Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holyoak S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders R., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volkart G., Aert R., Robben J., Gymnopoulos B.,
 RA Welfens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs W., Fritze C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Drees S., Gloux S., Lelarge V., Mottier S.,
 RA Galbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rocher M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe." ;
 RL Nature 415:871-880(2002).
 CC -I- FUNCTION: THE DISULFIDE BOND FUNCTIONS AS AN ELECTRON CARRIER IN
 CC THE GLUTATHIONE-DEPENDENT SYNTHESIS OF DEOXYRIBONUCLEOTIDES BY THE
 CC ENZYME RIBONUCLEOTIDE REDUCTASE. IN ADDITION, IT IS ALSO INVOLVED
 CC IN REDUCING SOME DISULFIDES IN A COUPLED SYSTEM WITH GLUTATHIONE
 CC REDUCTASE. THIOLESTERASE CATALYZES CELLULAR THIOLESTER
 CC TRANSFERASE REACTIONS. IT TRANSFERS REDUCING EQUIVALENTS
 CC TO CYTOSOLIC PROTEIN AND NONPROTEIN DISULFIDES (BY SIMILARITY).
 CC -I- SIMILARITY: BELONGS TO THE GLUTAREDOXIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z98980, CAB11722.1; -
 DR EMBL: AB015167, BAA28750.1; -
 DR EMBL: AF192764, AAF19628.1; -
 DR EMBL: AF121275, AAD25391.1; -
 DR PIR: T38824, T38824.
 DR HSSP: P35754, LTHR.
 DR GeneDB: SPombe, SPAC4F10.20; -
 DR InterPro: IPR002109, Glutaredoxin.
 DR InterPro: IPR006663, Thioledeox dom2.
 DR Pfam: PF00462, glutaredoxin; 1.
 DR PRINTS: PR00160, GLUTAREDOXIN.
 DR PROSITE: PS00195, GLUTAREDOXIN; 1.
 DR Redox-active center, Electron transport.
 KW DISULFID 25 REDOX-ACTIVE (BY SIMILARITY).
 FT CONFLICT 52 28 N -> D (IN REF. 3).

SQ SEQUENCE 101 AA; 11261 MW; 30557E19BF33E9BB CRC64;
 Query Match 0.9%; Score 7; DB 1; Length 101;
 Best local Similarity 100.0%; Pred. No. 26;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 461 NSDFQAL 467
 Db 82 NSDFQAL 88
 RESULT 16
 RS16 UREPA STANDARD; PRT; 101 AA.
 ID RS16 UREPA
 AC G9P5L1
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 30S ribosomal protein S16.
 GN RPS16 OR RPS16 OR U5568.
 OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.
 OX NCBI_TaxId=134621;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=serovar 3;
 RX MEDLINE=20500219; PubMed=11048724;
 RA Glass J.I., Letkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
 RA Cassell G.H.;
 RT "The complete sequence of the mucosal pathogen Ureaplasma
 RL urealyticum." ;
 RL Nature 407:757-762(2000).
 CC -I- SIMILARITY: BELONGS TO THE S16 FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE002155, AAF30982.1; -
 DR HSSP: P80379, 1EMW.
 DR HAMAP: MF_00385; -; 1.
 DR InterPro: IPR000307, Ribosomal_S16.
 DR Pfam: PF00886, Ribosomal_S16; 1.
 DR ProDom: PD003791, Ribosomal_S16; 1.
 DR TIGRFAMs: TIGR00002, S16; 1.
 DR PROSITE: PS00732, RIBOSOMAL_S16; FALSE_NEG.
 KW Ribosomal protein; Complete proteome
 SQ SEQUENCE 101 AA; 11297 MW; FED91E1AB36CBF45 CRC64;
 Query Match 0.9%; Score 7; DB 1; Length 101;
 Best local Similarity 100.0%; Pred. No. 26;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 404 PSETVKN 410
 Db 66 PSETVKN 72
 RESULT 17
 YQCC HAEIN STANDARD; PRT; 106 AA.
 ID YQCC HAEIN
 AC Q57152; Q05061;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein H11436.
 GN H11436.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;

```

CC Pasteurellaceae; Haemophilus.
OX NCBI_TaxId=727;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=rd / KM20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleisichman R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McInerney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utecherback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.,
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd."
RL Science 269:496-512(1995).
CC -1- SIMILARITY: TO THE N-TERMINAL OF E.CAROTENOVA EXOENZYM REULATION
CC REGULON ORFL. THE C-TERMINAL PART IS COLINEAR WITH YOCB.
CC -1- SIMILARITY: STRONG, TO E.COLI YOCB.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U32822; AAC23085.1; -
CC PIR: I64171; I64171.
CC TIGR: H11436; -
DR Pfam: PF04287; DUF446; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 106 AA; 12273 MW; 0955920EBD63228C CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 28; Length 106;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 207 AAEAFLS 213
DB 28 AAEAFLS 34

RESULT 18
RNPA STRAP STANDARD; PRT; 115 AA.
AC O8CMN4;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Ribonuclease P protein component (EC 3.1.26.5) (RNase protein) (RNase
DE P protein) (Protein C5).
GN RNPA OR SE2418.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxId=1282;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 12228;
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
RA Chen Z., Wen Y.;
RL Submitted (NOV-2002) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: RNase catalyzes the removal of the 5'-leader sequence
CC from pre-rRNA to produce the mature 5' terminus. It can also cleave
CC other RNA substrates such as 4.5S RNA. The protein component plays
CC an auxiliary but essential role in vivo by binding to the 5'-
CC leader sequence and broadening the substrate specificity of the
CC ribozyme (By similarity).
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage of RNA, removing 5'-
CC extra-nucleotide from tRNA precursor.

```

```

CC -1- SUBUNIT: Consists of a catalytic RNA component (M1 or rnpB) and a
CC protein subunit (By similarity).
CC -1- SIMILARITY: BELONGS TO THE RNPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE016752; AAC06061.1; -
CC HAMAP: MF_00227; -1.
DR Pfam: PF003629; Ribonuclease_P; 1.
DR Prodom: PD003629; Ribonuclease_P; 1.
DR TIGRFAMs: TIGR00188; rnpA; 1.
DR PROSITE: PS00648; RIBONUCLEASE_P; 1.
KW Hydrolase; Nuclease; Endonuclease; RNA processing; RNA-binding;
KW Complete proteome.
SQ SEQUENCE 115 AA; 13484 MW; 106B2592C8400F18 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 30; Length 115;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 460 RNSDFOA 466
DB 9 RNSDFOA 15

RESULT 19
INSC ECOLI STANDARD; PRT; 121 AA.
AC P19776; O07989; O08018; O08019; P76357; P77346;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transposase insc for insertion element IS2a/D/F/H/I/K.
GN (INSC OR B0160) AND (INSC OR B1403) AND (INSC OR B1997) AND
GN (INSC4 OR B2861) AND (INSC5 OR B3044) AND (INSC6 OR B4272).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxId=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=K12;
RX MEDLINE=88137965; PubMed=2830172;
RA Ronecker H.J., Rak B.;
RT "Genetic organization of insertion element IS2 based on a revised
RT nucleotide sequence."
RL Gene 59:291-296(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RX MEDLINE=9534362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes."
RL Nucleic Acids Res. 23:2105-2119(1995).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RX MEDLINE=9742617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RL "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [4]

```

RP SEQUENCE FROM N.A.
DT STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Alha H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,
RA Kitagawa M., Makino K., Miki T., Mizouchi K., Mori H., Mori T.,
RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,
RA Oshima T., Saito N., Sampel G., Seki Y., Sivasundaram S.,
RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,
RA Yamamoto Y., Horuchi T.;
RT "A 570-kb DNA sequence of the *Escherichia coli* K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
[5]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Alha H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,
RA Mizouchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y.,
RA Nishimoto H., Nishio Y., Oshima T., Saito N., Sampel G., Seki Y.,
RA Sivasundaram S., Tagami H., Takeda J., Takemoto K., Wada C.,
RA Yamamoto Y., Horuchi T.;
RT "A 460-kb DNA sequence of the *Escherichia coli* K-12 genome
corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
CC -!- FUNCTION: INVOLVED IN THE TRANSPOSITION OF THE INSERTION
SEQUENCE IS2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, V00279; CAA23542.1; -;
DR EMBL, U14003; AAA97168.1; ALT_INIT.
DR EMBL, U28377; AAA69212.1; ALT_INIT.
DR EMBL, U28375; AAA63043.1; ALT_INIT.
DR EMBL, AE000143; AAC73463.1; ALT_INIT.
DR EMBL, AE000237; AAC74485.1; ALT_INIT.
DR EMBL, AE000291; AAC75058.1; ALT_INIT.
DR EMBL, AE000369; AAC75900.1; -;
DR EMBL, AE000386; AAC76080.1; -;
DR EMBL, AE000498; AAC77228.1; ALT_INIT.
DR EMBL, D90778; BAA15013.1; ALT_INIT.
DR EMBL, D90779; BAA15019.1; ALT_INIT.
DR EMBL, D90838; BAA15822.1; ALT_INIT.
DR EMBL, D90850; BAA16005.1; ALT_INIT.
DR EMBL, D90851; BAA16013.1; ALT_INIT.
DR EMBL, D90852; BAA16036.1; ALT_INIT.
DR Ecocore; EG40003; insc.
DR InterPro; IPR002514; Transposase_8.
DR Pfam; PF01527; Transposase_8; 1.
KW Transposable element; Transposition; DNA-binding; DNA recombination;
KW Complete proteome.
FT VARIANT 34 34 L -> F (IN B1997).
SQ SEQUENCE 121 AA; 13452 MW; 59431B5C452E067A CRC64;
Query Match 0.9%; Score 7; DB 1; Length 121;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

AC P59444;
DT 15-SEP-2003 (Rel. 42; Created)
DT 15-SEP-2003 (Rel. 42; Last sequence update)
DT 15-SEP-2003 (Rel. 42; Last annotation update)
DE Transposase insc for insertion element IS2.
GN (INSC1 OR SF0245) AND (INSC2 OR SF0879) AND (INSC3 OR SF0933) AND
GN (INSC4 OR SF0960) AND (INSC5 OR SF1054) AND (INSC6 OR SF1165) AND
GN (INSC7 OR SF1443) AND (INSC8 OR SF1463) AND (INSC9 OR SF1587) AND
GN (INSC10 OR SF2011) AND (INSC11 OR SF2615) AND
GN (INSC12 OR SF2694) AND (INSC13 OR SF2873) AND
GN (INSC14 OR SF2984) AND (INSC15 OR SF3431) AND
GN (INSC16 OR SF3512) AND (INSC17 OR SF3805) AND
GN (INSC18 OR SF3873) AND (INSC19 OR SF3988) AND
GN (INSC20 OR SF4097) AND (INSC21 OR SF4185).
OS *Shigella flexneri*.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; *Shigella*.
OX NCBI_TaxID=623;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Xu J.;
RT "Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity
RT through comparison with genomes of *Escherichia coli* K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441(2002).
CC -!- FUNCTION: Involved in the transposition of the insertion
CC sequence IS2 (By similarity).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AE015060; AAM41906.1; ALT_INIT.
DR EMBL, AE015116; AAM42511.1; ALT_INIT.
DR EMBL, AE015123; AAM42562.1; ALT_INIT.
DR EMBL, AE015125; AAM42588.1; ALT_INIT.
DR EMBL, AE015133; AAM42676.1; ALT_INIT.
DR EMBL, AE015143; AAM42781.1; ALT_INIT.
DR EMBL, AE015160; AAM42947.1; ALT_INIT.
DR EMBL, AE015170; AAM43060.1; ALT_INIT.
DR EMBL, AE015180; AAM43173.1; ALT_INIT.
DR EMBL, AE015218; AAM43556.1; ALT_INIT.
DR EMBL, AE015277; AAM44112.1; ALT_INIT.
DR EMBL, AE015284; AAM44187.1; ALT_INIT.
DR EMBL, AE015301; AAM44359.1; ALT_INIT.
DR EMBL, AE015311; AAM44465.1; ALT_INIT.
DR EMBL, AE015351; AAM44892.1; ALT_INIT.
DR EMBL, AE015359; AAM44970.1; ALT_INIT.
DR EMBL, AE015387; AAM45245.1; ALT_INIT.
DR EMBL, AE015394; AAM45310.1; ALT_INIT.
DR EMBL, AE015405; AAM45422.1; ALT_INIT.
DR EMBL, AE015418; AAM45522.1; ALT_INIT.
DR EMBL, AE015427; AAM45606.1; ALT_INIT.
KW Transposable element; Transposition; DNA-binding; DNA recombination.
SQ SEQUENCE 121 AA; 13452 MW; 59431B5C452E067A CRC64;

Query Match 0.9%; Score 7; DB 1; Length 121;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 202 ASELAA 208
DB 70 ASELAA 76

RESULT 20
INSC_SHIFL
ID INSC_SHIFL STANDARD; PRT; 121 AA.

```

RESULT 21
RNP_ANTAM STANDARD; PRT; 124 AA.
ID RNP ANTAM
AC 200668;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).
GN RNASE1 OR RNS1.
OS Antilocapra americana (Pronghorn).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Antilocapridae; Antilocapra.
OX NCBI_TaxID=9891;
RN (1)
RP SEQUENCE.
RC TISSUE=Pancreas;
RX MEDLINE=80075014; PubMed=513141;
RA Beintema J.J., Gaastera W., Munnikema J.;
RT "Primary structure of pronghorn pancreatic ribonuclease: close
RL relationship between giraffe and pronghorn.";
CC J. Mol. Evol. 13:305-316(1979).
CC -1 CATALYTIC ACTIVITY: Endonucleolytic cleavage to 3'-
phosphates and 3'-phosphoglyconucleotides ending in C-P or U-P
with 2',3'-cyclic phosphate intermediates.
CC -1 SUBCELLULAR LOCATION: Secreted.
CC -1 TISSUE SPECIFICITY: Pancreas.
CC -1 SIMILARITY: Belongs to the pancreatic ribonuclease family.
DR PIR; A00813; NRPXH.
DR HSSP; P00656; IRBG.
DR InterPro; IPR001427; RNaseA.
DR Pfam; PF00074; RNaseA.1.
DR PRINTS; PR00794; RIBONUCLEASE.
DR PRODOM; PP000535; RNaseA.1.
DR SMART; SM00092; RNase_Pc.1.
DR PROSITE; PS00127; RNase_PANCREATIC; 1.
KW Hydrolase; Nuclease; Endonuclease; Glycoprotein.
FT DISULFID 26 84 BY SIMILARITY.
FT DISULFID 40 95 BY SIMILARITY.
FT DISULFID 58 110 BY SIMILARITY.
FT DISULFID 65 72 BY SIMILARITY.
FT ACT_SITE 12 12 BY SIMILARITY.
FT ACT_SITE 41 41 BY SIMILARITY.
FT ACT_SITE 119 119 BY SIMILARITY.
FT CARBOHYD 34 34 N-LINKED (GLCNAC... ) (PARTIAL).
SQ SEQUENCE 124 AA; 13711 MW; 9435EF532420F852 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 785 SNPSVS 791
DB 15 SNPSVS 21

RESULT 22
Y082_RICPR STANDARD; PRT; 143 AA.
ID Y082_RICPR
AC Q92E65;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein RP082.
GN RP082.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=782;
RN (1)
RP SEQUENCE FROM N.A.
RX STRAIN=Madrid E;

```

```

RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sichenitz-Ponten T., Alsmark U.C.M., Podowski R.M., Naestlund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RL mitochondria.";
RL Nature 396:133-140(1998).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ235270; CAA14552.1; -
DR PIR; A71717; A71717.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 143 AA; 16939 MW; 57637C7A62FE9F10 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 143;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 NRKOEI 139
DB 106 NRKOEI 112

RESULT 23
P7VA_ECOLI STANDARD; PRT; 148 AA.
ID P7VA_ECOLI
AC P3155; P76776;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE PTS system, fructose-1,6-bisphosphate (Phosphotransferase enzyme
DE II, A component) (EC 2.7.1.69).
CN PFVA OR B3900.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=93374854; PubMed=8396120;
RA Moralejo P., Egan S.M., Hidalgo B.F., Aguilar J.;
RT "Sequencing and characterization of a gene cluster encoding the
RT enzymes for L-thiamine metabolism in Escherichia coli.";
RL J. Bacteriol. 175:5585-5594(1993).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=93347969; PubMed=8346018;
RA Plunkett G. III, Burland V.D., Daniels D.L., Blattner F.R.;
RT "Analysis of the Escherichia coli genome. III. DNA sequence of the
RT region from 87.2 to 89.2 minutes.";
RL Nucleic Acids Res. 21:3391-3398(1993).
RN (3)
RP REVISIONS TO 81 AND 104-108.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN (4)
RP DISCUSSION OF SEQUENCE.
RX MEDLINE=94290319; PubMed=8019415;

```

RA Reizer J., Micholey V., Reizer A., Saier M.H. Jr.;
 RT "Novel phosphotransferase system genes revealed by bacterial genome
 RT analysis: unique, putative fructose- and glucoside-specific
 RT systems.";

RL Protein Sci. 3:440-450(1994)

CC -1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
 CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
 CC TRANSPORT SYSTEM. THE IIDC DOMAIN CONTAINS THE SUGAR BINDING SITE
 CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
 CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-PHF); IIA TRANSFERS ITS
 CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
 CC THE SUGAR.

CC -1- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
 CC histidine + sugar phosphate.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).

CC -1- SIMILARITY: Contains 1 PTS EIIA domain.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X60472; CAA43004.1; -
 DR EMBL; L19201; AAB0303.2; -
 DR EMBL; AE000465; AAC76882.1; -
 DR PIR; D48649; D48649.
 DR EcoGene; EG1864; ftyA.
 DR InterPro; IPR002178; PTS_EIIA_2.
 DR InterPro; IPR004715; PTSIIA_fnc.
 DR Pfam; PF00359; PTS_EIIA_2; 1.
 DR ProDom; PD001689; PTS_EIIA_2; 1.
 DR TIGRfam; TIGR00848; ftnA_1.
 DR PROSITE; PS00372; PTS_EIIA_2; 1.
 DR Phosphotransferase system; Sugar transport; Transferase;
 KW Phosphorylation; Complete proteome.

FT MOD RES 64 64 PHOSPHORYLATION (BY SIMILARITY).
 FT CONFLICT 81 81 A -> G (IN REF. 2).
 FT CONFLICT 104 107 QSGE -> KXZ (IN REF. 2).
 SQ SEQUENCE 148 AA; 16093 MW; 04AE87B9084BA1C6 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 148;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 303 GVAVPHG 309
 Db 59 GVAVPHG 65

RESULT 24
 ID MUTT STRAM STANDARD; PRT; 154 AA.
 AC P32091;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Mutt-like protein (ORF154).
 OS Streptomyces ambifaciens.
 OC Plasmid pSAM2.
 CC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 CC Streptomycinae; Streptomycetaceae; Streptomycetes.
 OX NCBI_TaxID=1889;
 RN [1]
 RP .SEQUENCE FROM N.A.
 RC STRAIN=ATCC 23877;
 RX MEDLINE=95020551; PubMed=7934842;
 RA Haggag U., Bernode U.L., Friedmann A., Guerin M.;
 RT "Mode and origin of replication of pSAM2, a conjugative integrating
 RT element of Streptomyces ambifaciens.";
 RL Mol. Microbiol. 10:799-812(1993).

CC -1- SIMILARITY: BELONGS TO THE NUDIX HYDROLASE FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; Z19590; CAA79638.1; -
 DR PIR; S39873; S39873.
 DR HSSP; P08337; 1TUM.
 DR InterPro; IPR000086; NUDIX_hydrolase.
 DR Pfam; PF00293; NUDIX; 1.
 DR PRINTS; PR00502; NUDIXFAMILY.
 DR PROSITE; PS00893; NUDIX; 1.
 KW Plasmid; Hydrolase.
 FT DOMAIN 48 69 NUDIX BOX.
 SQ SEQUENCE 154 AA; 16689 MW; 8EC2EE99D272D656 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 154;
 Best Local Similarity 100.0%; Pred. No. 38;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 LTPDEV3 39
 Db 114 LTPDEV3 120

RESULT 25
 ID GREX STAM STANDARD; PRT; 158 AA.
 AC O99TN9;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Transcription elongation factor greA (Transcript cleavage factor
 DE greA).
 GN GREX OR SA11610 OR SA1438 OR MW1560.
 OS Staphylococcus aureus (strain M50 / ATCC 700699),
 OS Staphylococcus aureus (strain N315), and
 OS Staphylococcus aureus (strain MW2).
 CC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=15878, 158879, 196620;
 RN [1]
 RP .SEQUENCE FROM N.A.
 RC STRAIN=M50, and N315;
 RX MEDLINE=21311952; PubMed=11418146;
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
 RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-O., Ito T.,
 RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
 RA Mizutani U. Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
 RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
 RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
 RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus
 RT aureus.";
 RL Lancet 357:1225-1240(2001).
 RN [2]
 RP .SEQUENCE FROM N.A.
 RC STRAIN=MW2;
 RX MEDLINE=22040717; PubMed=12044378;
 RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
 RA Nagai Y., Iwama N., Asano K., Naim T., Kuroda H., Cui L.,
 RA Yamamoto K., Hiramatsu K.;
 RT "Genome and virulence determinants of high virulence community-
 RT acquired MRSA.";
 RL Lancet 359:1819-1827(2002).

CC -1- FUNCTION: Necessary for efficient RNA polymerase transcription
 CC elongation past template-encoded arresting sites. The arresting
 CC sites in DNA have the property of tripping a certain fraction of
 CC elongating RNA polymerases that pass through, resulting in locked

```

CC ternary complexes. Cleavage of the nascent transcript by cleavage
CC factors such as grea or greb allows the resumption of elongation
CC from the new 3' terminus. Grea releases sequences of 2 to 3
CC nucleotides (By similarity).
CC -1- SIMILARITY: BELONGS TO THE GREB/GREB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP003362; BAB5772.1; -
DR EMBL; AP003334; BAB42702.1; -
DR EMBL; AP004827; BAB95425.1; -
DR PIR; A89943; A89943.
DR HSSP; P21346; IGRJ.
DR HAMAP; MF_00105; -; 1.
DR InterPro; IPR006359; Grea.
DR InterPro; IPR001437; Grea_Greb.
DR Pfam; PF01272; Grea_Greb; 1.
DR Pfam; PF03449; Grea_Greb_N; 1.
DR ProDom; PD004918; Grea_Greb; 1.
DR TIGRFAMs; TIGR01462; grea; 1.
DR PROSITE; PS00829; GREAB_1; 1.
DR PROSITE; PS00830; GREAB_2; 1.
DR PROSITE; PS00830; GREAB_2; 1.
DR Transcription regulation; DNA-binding; Coiled coil; Complete proteome.
DR DOMAIN 4 COILED COIL (POTENTIAL).
DR SEQUENCE 158 AA; 17743 MW; EC3B0F0E6238A107 CRC64;
SQ
Query Match 0.9%; Score 7; DB 1; Length 158;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 176 IIEDTGD 182
DB 77 IIEDTGD 83
RESULT 26
Y052_ARCFU STANDARD; PRT; 158 AA.
AC C30184.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AF0052.
GN AF0052.
OS Archaeoglobus fulgidus.
OC Archaeae; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=96049343; PubMed=9389475;
RA Kleink H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kesteven A.R., Graham D.E., Kyrylova N.C.,
RA Fleischman R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kitzner S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Peterson S., Gocayne J.D., Weidman J.F., McDonald L., Uterback T.,
RA Cocton P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -----

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001102; AAB91174.1; -
DR PIR; D69256; D69256.
DR TIGR; AF0052; -
DR Hypothetical protein; Complete proteome.
DR SEQUENCE 158 AA; 17369 MW; 92DF7317D55ABFC CRC64;
SQ
Query Match 0.9%; Score 7; DB 1; Length 158;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 372 NSSLVSQ 378
DB 36 NSSLVSQ 42
RESULT 27
GCSI_ARATH STANDARD; PRT; 166 AA.
ID GCSI_ARATH
AC O91010.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable glycine cleavage system H protein 2, mitochondrial precursor.
GN ATIG32470 OR F5D14.25.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk D.B., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Creasy T.H., Dewar K.,
RA Dunn P., Egu P., Feldhahn T.V., Feng J.-D., Fong B., Fujii C.I.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hutzar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Lueros J.S., Maiti R., Marzilli A.,
RA Miltseher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana."
RL Nature 408:816-820(2000).
CC -1- FUNCTION: THE GLYCINE CLEAVAGE SYSTEM CATALYZES THE DEGRADATION OF
CC GLYCINE. THE H PROTEIN SHUTTLES THE METHYLAMINE GROUP OF GLYCINE
CC FROM THE P PROTEIN TO THE T PROTEIN (BY SIMILARITY).
CC -1- CORACOR: THE H CHAIN CONTAINS A COVALENTLY-BOUND LIPOYL COFACTOR
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Mitochondrial (By similarity).
CC -1- SIMILARITY: Contains 1 lipoyl-binding domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```


CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AC007767; AAF81345.1; -
DR PIR; A86450; A86450.
DR HSSP; P16048; 1HTP.
DR InterPro; IPR002930; GCV_H.
DR InterPro; IPR003016; LipoyL.
DR Pfam; PF01597; GCV_H; 1.
DR TIGRfam; TIGR00527; GCVH; 1.
DR PROSITE; PS00189; LipoyL; 1.
KW Mitochondrion; Transist peptide; LipoyL.
FT TRANSIT 1 35 MITOCHONDRION (BY SIMILARITY).
FT CHAIN 36 166 PROBABLE GLYCINE CLEAVAGE SYSTEM H
FT BINDING 98 98 LipoYL (BY SIMILARITY).
FT SEQUENCE 166 AA; 17897 MW; C21658C0B9E75ADB CRC64;

Query Match 0.9%; Score 7; DB 1; Length 166;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 788 SSVSKER 794
Db 83 SSVSKER 89

RESULT 28
HUNB DROMU STANDARD; PRT; 174 AA.
AC 046250; 046251;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hunchback protein (Fragments).
OS HB.
GN Drosophila mulleri (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7231;
RN [1]
RP SEQUENCE FROM N.A.
RA Baker R.H., Desalle R.;
RT "Multiple sources of character information and the phylogeny of
RT Hawaiian Drosophilids.";
RL Syst. Biol. 46:654-673(1997).
CC -1- FUNCTION: GAP CLASS SEGMENTATION PROTEIN THAT CONTROLS DEVELOPMENT
CC OF HEAD STRUCTURES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear (by similarity).
CC -1- SIMILARITY: BELONGS TO THE HUNCHBACK FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U93014; AAC03262.1; -
DR EMBL; U93015; AAC03263.1; -
DR FlyBase; FBgn0023760; Dmull.hb.
KW Developmental protein; Gap protein; Zinc-finger;
KW Metal-binding; DNA-binding; Repeat; Nuclear protein.
FT NON_TER 1 1
FT DOMAIN 58 67 POLY-GLN.
FT NON_CONS 93 94
FT NON_TER 174 174
SQ SEQUENCE 174 AA; 19265 MW; DA27CF6CC8CC8368 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 174;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 349 PFPQPTP 355
Db 110 PFPQPTP 116

RESULT 29
APT CHLITE STANDARD; PRT; 177 AA.
AC 08KFM9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Adenine phosphoribosyltransferase (EC 2.4.2.7) (APRT).
GN APT OR C70293.
OS Chlorobium tepidum.
OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
OC Chlorobium.
OX NCBI_TaxID=1097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TLS / ATCC 49652 / DSM 12025;
RX MEDLINE=22103685; PubMed=12093901;
RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,
RA Dodson R.J., Deboy R., Gwim M.L., Nelson W.C., Hatt D.H.,
RA Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,
RA Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,
RA Nierman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Radune D.,
RA Vamathavan J., Khouri H., White O., Gruber T.M., Ketchum K.A.,
RA Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;
RT "The complete genome sequence of Chlorobium tepidum TLS, a
RT photosynthetic, anaerobic, green-sulfur bacterium.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).
CC -1- FUNCTION: Catalyzes a salvage reaction resulting in the formation
CC of AMP, that is energetically less costly than de novo synthesis.
CC -1- CATALYTIC ACTIVITY: AMP + diphosphate = adenine + 5-phospho-
CC alpha-D-ribose 1-diphosphate.
CC -1- PATHWAY: Purine salvage.
CC -1- SUBUNIT: Homodimer (by similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE PURINE/PYRIMIDINE
CC PHOSPHORIBOSYLTRANSFERASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AE012808; AAM71539.1; -
DR TIGR; C70293; -
DR HAMAP; MF_00004; -; 1.
DR InterPro; IPR005764; Ade_phospho_trans.
DR InterPro; IPR002375; Pr/py rp transf.
DR InterPro; IPR006836; PRTtransferase.
DR Pfam; PF00156; Pribosyltran; 1.
DR TIGRfam; TIGR01090; apt; 1.
DR PROSITE; PS00103; PUR_PYP_TRANSER; 1.
KW Transferase; Glycosyltransferase; Purine salvage; Complete proteome.
SQ SEQUENCE 177 AA; 19319 MW; A060DD2DD9C242B41 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 177;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 484 LVDDLLA 490

DB 118 LVDDL1A 124

RESULT 30

NUSG_NEIMA STANDARD; PRT; 178 AA.

ID NUSG_NEIMA

AC Q9JRD9;

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Transcription antitermination protein nusG.

GN NUSG OR NMA0147 OR NMB0126.

OS Neisseria meningitidis (serogroup A), and

OC Neisseria meningitidis (serogroup B).

OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;

OC Neisseriaceae; Neisseria.

OX NCBI_TaxID=65699, 491;

RP SEQUENCE FROM N.A.

RC STRAIN=Z2491 / Serogroup A / Serotype 4A;

RX MEDLINE=20222556; PubMed=10761919;

RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C., Klee S.R., Mottell G., Basham D., Brown D., Chillingworth T., Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holtroyd S., Jagels K., Leather S., Moule S., Mungall K., Quail M.A., Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J., Whitehead S., Spratt B.G., Barrall B.G.,

RT "Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491."

RL Nature 404:502-506(2000).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=MC58 / Serogroup B;

RX MEDLINE=20175751; PubMed=10710307;

RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E., Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J., Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K., Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A., Mason T., Ciecko A., Parksey D.S., Blair E., Clifton H., Clark E.B., Cotton M.D., Ulterback T.R., Khouri H., Qin H., Vamathevan J., Gill J., Scarlato V., Maignan V., Pizza M., Grandi G., Sun L., Smith H.O., Fraser C.M., Moxon E.R., Rappunli R., Venter J.C.,

RT "Complete genome sequence of Neisseria meningitidis serogroup B strain MC58."

RL Science 287:1809-1815(2000).

CC -1- FUNCTION: INFLUENCES TRANSCRIPTION TERMINATION AND ANTI-TERMINATION. ACTS AS A COMPONENT OF THE TRANSCRIPTION COMPLEX, AND INTERACTS WITH THE TERMINATION FACTOR RHO AND RNA POLYMERASE (BY SIMILARITY).

CC -1- SIMILARITY: Belongs to the nusG family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AL162752; CAB83462.1; -

DR EMBL; AE002371; AAP40585.1; -

DR PIR; C81235; C81235.

DR TIGR; NMB0126; -

DR InterPro; IPR005824; K0M.

DR InterPro; IPR006646; K0M_sub.

DR InterPro; IPR006645; NGN.

DR InterPro; IPR001062; NusG.

DR Pfam; PF00467; K0M; 1.

DR Pfam; PF02357; NusG; 1.

DR PRINTS; PR00338; NUSGINSCTPCT.

DR SMART; SM00739; K0M; 1.

DR SMART; SM00738; NGN; 1.

DR TIGRFAMs; TIGR00922; nusG; 1.

DR PROSITE; PS01014; NUSG; 1.

KW Transcription termination; Complete proteome.

SC SEQUENCE 178 AA; 20550 MW; 3171FDC957EFCB3 CRC64;

Query Match

Best local similarity 100.0%; Score 7; DB 1; Length 178;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 323 LEER1AR 329

DB 22 LEER1AR 28

RESULT 31

APT_HELPJ STANDARD; PRT; 179 AA.

ID APT_HELPJ

AC Q9ZIQ9;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Adenine phosphoribosyltransferase (EC 2.4.2.7) (APRT).

GN APT OR JHP0519.

OS Helicobacter pylori J99 (Campylobacter pylori J99).

OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;

OC Helicobacteraceae; Helicobacter.

OX NCBI_TaxID=85963;

RP SEQUENCE FROM N.A.

RX MEDLINE=99120557; PubMed=9923682;

RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C., Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G., Tummino P.J., Carnuso A., Uria-Nickelsen M., Mills D.M., Ives C., Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F., Trust T.J.;

RT "Genomic sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori."

RL Nature 397:176-180(1999).

CC -1- FUNCTION: Catalyzes a salvage reaction resulting in the formation of AMP, that is energetically less costly than de novo synthesis.

CC -1- CATALYTIC ACTIVITY: AMP + diphosphate = adenine + 5-phospho-alpha-D-ribose 1-diphosphate.

CC -1- PATHWAY: Purine salvage.

CC -1- SUBUNIT: Homodimer (By similarity).

CC -1- SUBCELLULAR LOCATION: Cytoplasmic.

CC -1- SIMILARITY: BELONGS TO THE PURINE/PRIMIDINE PHOSPHORIBOSYLTRANSFERASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AE001485; AAD06100.1; -

DR PIR; H71920; H71920.

DR HAMAP; MF_00004; -; 1.

DR InterPro; IPR005764; Ade_phospho_trans.

DR InterPro; IPR002375; Pr/py_rp_transf.

DR InterPro; IPR000836; PRTnsfrase.

DR Pfam; PF00156; Pribosyltran; 1.

DR TIGRFAMs; TIGR01090; apt; 1.

DR PROSITE; PS00103; PUR_PYR_PR_TRANSFR; 1.

KW Transferase; Glycosyltransferase; Purine salvage; Complete proteome.

SC SEQUENCE 179 AA; 19846 MW; F287FDAA23CB57C9 CRC64;

Query Match

Best local similarity 100.0%; Score 7; DB 1; Length 179;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 484 LVDDL1A 490

|||||

DB 124 LVDDLLA 130

RESULT 32

VATE DEIRA STANDARD; PRT; 185 AA.

AC Q9RWH1; 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE V-type ATP synthase subunit E (EC 3.6.3.14) (V-type ATPase subunit E).

GN ATP8 OR DR0697.

OS Deinococcus radiodurans.

OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;

OC Deinococcaceae; Deinococcus.

OX NCBI_Taxid=1299;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;

RX MEDLINE=20036896; PubMed=10567266;

RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,

RA Dodson R.J., Haft D.H., Gwin M.L., Nelson W.C., Richardson D.L.,

RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,

RA Vamathevan J.U., Lam P., McDonald L., Utterback T., Zalewski C.,

RA Makarewicz K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,

RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,

RA Fraser C.M.;

RA "Genome sequence of the radioresistant bacterium Deinococcus

RT radiodurans R1.";

RL Science 286:1571-1577(1999).

CC -1- FUNCTION: Produces ATP from ADP in the presence of a proton

CC gradient across the membrane.

CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +

CC H(+) (Out).

CC -1- SIMILARITY: Belongs to the V-ATPase E subunit family.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

CC EMBL; AE001926; AAF10275.1; -

DR PIR; F75487; F75487.

DR TIGR; DR0697; -

DR HAMAP; MF_00311; -; 1.

DR InterPro; IPR002842; ATPsyn Esub.

DR Pfam; PF01991; VATP-synt E; 1.

KW Hydrolase; ATP synthetase; Hydrogen ion transport;

KW Complete proteome.

SQ SEQUENCE 185 AA; 20140 MW; CF4F2F8987C3CF2 CRC64;

QY 466 ALDKLIE 472

DB 2 ALDKLIE 8

Query Match 0.9%; Score 7; DB 1; Length 185;

Best Local Similarity 100.0%; Pred. No. 45;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Methylobacteriaceae; Methylobacterium.

OX NCBI_Taxid=408;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=AM1 / NCIMB 9133;

RX MEDLINE=88319960; PubMed=2842733;

RA Nunn D.N., Anthony C.;

RA "The nucleotide sequence and deduced amino acid sequence of the genes

RT for cytochrome c1 and a hypothetical second subunit of the methanol

RT dehydrogenase of Methylobacterium AM1.";

RL Nucleic Acids Res. 16:7722-7722(1988).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=AM1 / NCIMB 9133;

RX MEDLINE=89134152; PubMed=2851998;

RA Nunn D.N., Anthony C.;

RA "The nucleotide sequence and deduced amino acid sequence of the

RT cytochrome c1 gene of Methylobacterium extorquens AM1, a novel class

RT of c-type cytochrome.";

RL Biochem. J. 256:673-676(1988).

RN [3]

RP SEQUENCE OF 1-8 FROM N.A.

RC STRAIN=AM1 / NCIMB 9133;

RA Anderson D.J., Morris C.J., Nunn D.N., Anthony C., Lidstrom M.E.;

RL Submitted (JAN-1990) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: ELECTRON ACCEPTOR FOR MDH. ACTS IN METHANOL OXIDATION.

CC THIS CYTOCHROME HAS A REDOX POTENTIAL OF ABOUT +256 MV.

CC -1- SUBCELLULAR LOCATION: Periplasmic (Potential).

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

CC EMBL; X07856; CAA30704.1; -

DR EMBL; M31108; AAA25382.1; -

DR HSSP; P29899; 2MTA.

DR InterPro; IPR003088; Cyt CI.

DR InterPro; IPR000345; Cytochrome c1.

DR Pfam; PF00034; Cytochrome c1.

DR PROSITE; PS00190; CYTOCHROME C1.

KW Electron transport; Heme; Signal; Methanol utilization; Periplasmic.

KW SIGNAL

FT CHAIN 1 25

FT BINDING 26 197

FT BINDING 90 90

FT BINDING 93 93

FT METAL 94 94

FT METAL ? ?

SQ SEQUENCE 197 AA; 21226 MW; 58F22BACB90258CE CRC64;

QY 31 ENLTPDE 37

DB 139 ENLTPDE 145

Query Match 0.9%; Score 7; DB 1; Length 197;

Best Local Similarity 100.0%; Pred. No. 48;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 34

RBIH HUMAN STANDARD; PRT; 201 AA.

AC Q9H0T4; 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Ras-related protein Rab-1B.

GN RAB1B.

OS Homo sapiens (Human).

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=21154917; PubMed=11230166;
RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
RA Anorge W., Boecker M., Bloeker H., Bauersachs S., Blum H.,
RA Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,
RA Mewes H.-W., Oltenswaelder B., Odetmaier B., Tampe J., Heubner D.,
RA Wambutt R., Korn B., Klein M., Poustka A.;
RA "Towards a catalog of human genes and proteins: sequencing and
RT analysis of 500 novel complete protein coding human cDNAs.";
RL Genome Res. 11:422-435(2001).
CC -!- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR
CC TRAFFIC (BY SIMILARITY).
CC -!- MISCELLANEOUS: RAB-1B BINDS GTP AND GDP AND POSSESS INTRINSIC
CC GTPASE ACTIVITY (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RAB FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AL136635; CAB66570.1; -.
CC DR Genew; HGNC:18370; RAB1B.
CC DR HSSP; P05713; 3RAB.
CC DR InterPro; IPR003579; GTPase_Rab.
CC DR InterPro; IPR001806; Ras_trnfsmg.
CC DR Pfam; PF00071; ras; 1.
CC DR PRINTS; PR00449; RASTRNSFRMNG.
CC DR SMART; SM00175; RAB; 1.
CC DR TIGRfam; TIGR00231; small GTP; 1.
CC KW GTP-binding; Lipoprotein; Prenylation; Protein transport.
CC FT NP_BIND 15 22 GTP (BY SIMILARITY).
CC FT NP_BIND 63 67 GTP (BY SIMILARITY).
CC FT NP_BIND 121 124 GTP (BY SIMILARITY).
CC FT DOMAIN 37 45 EFFECTOR REGION (BY SIMILARITY).
CC FT LIPID 200 200 GERANYL-GERANYL (BY SIMILARITY).
CC FT LIPID 201 201 GERANYL-GERANYL (BY SIMILARITY).
CC FT LIPID 201 201 GERANYL-GERANYL (BY SIMILARITY).
CC SQ SEQUENCE 201 AA; 22171 MW; 9812F4DAC34B2BE CRC64;

Query Match 0.9%; Score 7; DB 1; Length 201;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 365 PNLKIDS 371
DB 184 PNLKIDS 190

RESULT 35
RIB_RAT STANDARD; PRT; 201 AA.
AC P10536;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ras-related protein Rab-1B.
GN RAB1B.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89160341; PubMed=2493636;

```

```

RA Zahraoui A., Touchot N., Chardin P., Tavittian A.;
RT "Nucleotide sequence of a rat cDNA: rab1B, encoding a rab1-YPT
RT related protein.";
RL Nucleic Acids Res. 17:1770-1770(1989).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=9003316; PubMed=2509243;
RA Touchot N., Zahraoui A., Viel E., Tavittian A.;
RT "Biochemical properties of the YPT-related rab1B protein. Comparison
RT with rab1A."
RL FEBS Lett. 256:79-84(1989).
RN [3]
RP ISOPRENOLD.
RX MEDLINE=91296801; PubMed=1648736;
RA Khosravi-Far R., Lutz R.J., Cox A.D., Conroy L., Bourne J.R.,
RA Sienkney M., Balch W.E., Buss J.E., Der C.J.;
RT "Isoprenoid modification of rab proteins terminating in CC or CXC
RT motifs."
RL Proc. Natl. Acad. Sci. U.S.A. 88:6264-6268(1991).
CC -!- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR
CC TRAFFIC (BY SIMILARITY).
CC -!- MISCELLANEOUS: RAB-1B BINDS GTP AND GDP AND POSSESS INTRINSIC
CC GTPASE ACTIVITY.
CC -!- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RAB FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; X13905; CAA32105.1; -.
CC DR PIR; S06147; S06147.
CC DR HSSP; P05713; 3RAB.
CC DR InterPro; IPR003579; GTPase_Rab.
CC DR InterPro; IPR001806; Ras_trnfsmg.
CC DR InterPro; IPR005225; Small_GTP.
CC DR Pfam; PF00071; ras; 1.
CC DR PRINTS; PR00449; RASTRNSFRMNG.
CC DR SMART; SM00175; RAB; 1.
CC DR TIGRfam; TIGR00231; small GTP; 1.
CC KW GTP-binding; Lipoprotein; Prenylation; Protein transport.
CC FT NP_BIND 15 22 GTP (BY SIMILARITY).
CC FT NP_BIND 63 67 GTP (BY SIMILARITY).
CC FT NP_BIND 121 124 GTP (BY SIMILARITY).
CC FT DOMAIN 37 45 EFFECTOR REGION (BY SIMILARITY).
CC FT LIPID 200 200 GERANYL-GERANYL.
CC FT LIPID 201 201 GERANYL-GERANYL.
CC FT MUTAGEN 21 21 K->M; ABOLISHES GTP-BINDING.
CC SQ SEQUENCE 201 AA; 22163 MW; 8D3EDDC2AF4A2FE CRC64;

Query Match 0.9%; Score 7; DB 1; Length 201;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 365 PNLKIDS 371
DB 184 PNLKIDS 190

RESULT 36
APSI_SCHPO STANDARD; PRT; 210 AA.
AC Q09790;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Diadenosine 5',5'-P1,P6-hexaphosphate hydrolase (EC 3.6.1.-)
DE (Ap6A hydrolase).
GN APS1 OR SPAC13G6.14 OR SPAC24B11.03.
OX Schizosaccharomyces pombe (Fission yeast).
OS

```

Query Match 0.9%; Score 7; DB 1; Length 210;

```

Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      773 AEAERKL 779
       |||||
Db      143 AEAERKL 149

RESULT 37
14P_ARATH STANDARD; PRT; 211 AA.
AC 064628:
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protein At2g18990.
GN AT2G18990 OR F19F24.19.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxId=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Bairstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ranning C.M., Koo H.L.,
RA Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,
RA Tallon L.J., Gill J.E., Adams M.D., Carreira A.J., Creasy T.H.,
RA Goodman H.M., Somerville C.R., Copenhagen G.P., Preuss D.,
RA Mermer W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RL thaliana.";
RL Nature 402:761-768(1999).
CC -I- SIMILARITY: BELONGS TO THE UPF0071 FAMILY. SOME SIMILARITY TO THE
CC RHOREDOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on lists
CC use by non-profit institutions as long as their content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; AC003673; AAM14890.1; -.
DR PIR; T01627; T01627.
DR InterPro; IPR006663; Thioredox dom2.
SQ SEQUENCE 211 AA; 24451 MW; _A3B26F01CF7E6D14 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 211;
Best local similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      323 LEERIAR 329
       |||||
Db      169 LEERIAR 175

RESULT 38
14P_ECOLI STANDARD; PRT; 211 AA.
AC P45463;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yzar.
DE YZAR OR B3152.
GN Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

```

```

CC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Valdes J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC -1- SIMILARITY: TO YEAST YER004W.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U18997; AA57955.1; ALT_INIT.
CC EMBL; AE000396; AAC76186.1; ALT_INIT.
CC EcoGene; EG12783; YRAE.
CC KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 211 AA; 23198 MW; 5E3793BBADC891FF CRC64;

Query Match 0.9%; Score 7; DB 1; Length 213;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 600 YNRVKG 606
Db 122 YNRVKG 128

RESULT 39
CPCF_SYNEL STANDARD; PRT; 213 AA.
ID CPCF_SYNEL
AC P50038;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-EB-2003 (Rel. 41, Last annotation update)
DE Phycocyanobilin lyase beta subunit (Ec 4.1.1.1) (Phycocyanin operon
DE protein cpcF).
GN CPCF OR TLR1962.
OS Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RA Shimazu T., Soga M., Hirano M., Katoh S.;
RT "Cloning and sequencing of the phycocyanin operon from the
RT thermophilic cyanobacterium Synechococcus elongatus."
RL Submitted (SEP-1992) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=BP-1;
RX MEDLINE=2225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Matanabe A., Iriuchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1."
RL DNA Res. 9:123-130(2002).
CC -1- FUNCTION: REQUIRED FOR THE CHROMOPHYLLATION OF THE CP24 GENE
CC PRODUCT (BY SIMILARITY).
CC -1- SUBUNIT: CPCF AND CPCF ASSOCIATES TO FORM A LYASE.
CC -1- SIMILARITY: BELONGS TO THE CPCF/RPCF/BECP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D13173; BAA02460.1; -.
CC EMBL; AP005375; BAC09514.1; -.
DR InterPro; IPR004155; PBS_Lyase_HEAT.
DR Pfam; PF03130; HEAT_PBS; 3.
KW Phycobiosome; Lyase; Complete proteome.
SQ SEQUENCE 213 AA; 22654 MW; 45D99720985F2683 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 213;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 728 EAQLKEA 734
Db 135 EAQLKEA 141

RESULT 40
URE1_HELMU STANDARD; PRT; 213 AA.
ID URE1_HELMU
AC P50044;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Urease alpha subunit (EC 3.5.1.5) (Urea amidohydrolase) (Fragment).
GN UREA.
OS Helicobacter mustelae.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=217;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=NCCTC 12032;
RX MEDLINE=95369940; PubMed=7642313;
RA Solnick J.V., Josephans C., Tompkins L.S., Labigne A.;
RT "Construction and characterization of an isogenic urease-negative
RT mutant of Helicobacter mustelae."
RL Infect. Immun. 63:3718-3721(1995).
CC -1- CATALYTIC ACTIVITY: Urea + H(2)O = CO(2) + 2 NH(3).
CC -1- SUBUNIT: CONSIST OF TWO SUBUNITS (ALPHA AND BETA).
CC -1- CAUTION: IN HELICOBACTER THE ALPHA SUBUNIT IS WHAT IS KNOWN, IN
CC OTHER BACTERIA AS THE BETA SUBUNIT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L33462; AAC1483.1; -.
CC HSSP; P41022; IUBP.
DR InterPro; IPR002019; Urease_beta.
DR InterPro; IPR002026; Urease_gamma.
DR Pfam; PF00699; Urease_beta.1.
DR Pfam; PF00547; Urease_gamma.1.
DR PIRSF; PIRSF01225; Urease_gammadelta.1.
DR ProDom; PD002326; Urease_beta.1.
DR ProDom; PD002319; Urease_gamma.1.
DR TIGRFAMs; TIGR00192; Urease_beta.1.
DR TIGRFAMs; TIGR00193; Urease_gamma.1.
KW Hydrolase.
FT NON_TER 1
SQ SEQUENCE 213 AA; 23794 MW; B97A6A3FD2CB2F40 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 213;

```

Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 92 LKEDIV 98
Db 101 LKEDIV 107

RESULT 41

PSE2_RAT ID_PSE2_RAT STANDARD; PRT; 238 AA.

AC 063798;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Proteasome activator complex subunit 2 (proteasome activator 28-beta subunit) (PA28beta) (PA28b) (Activator of multicatalytic protease subunit 2) (11S regulator complex beta subunit) (REG-beta).
DE subunit 2) (11S regulator complex beta subunit) (REG-beta).
GN PSME2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxId=10116;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=95309399; PubMed=7789512;
RA Ahn J.Y., Tanabashi N., Akiyama K., Hisamatsu H., Noda C., Tanaka K., Chung C.H., Shibamura N., Willy P.J., Mott J.D., Slaughter C.A., Demetrio G.N.;
RA Demetrio G.N.;
RT "Primary structures of two homologous subunits of PA28, a gamma-interferon-inducible protein activator of the 20S proteasome.";
RL FEBS Lett. 366:37-42(1995).

CC -1- FUNCTION: Implicated in immunoproteasome assembly and required for efficient antigen processing. The PA28 activator complex enhances the generation of class I binding peptides by altering the cleavage pattern of the proteasome.
CC -1- SUBUNIT: HETERODIMER OF PSME1 AND PSME2, WHICH FORMS A HEXADIMERIC RING.
CC -1- INDUCTION: By interferon gamma.
CC -1- SIMILARITY: BELONGS TO THE PA28 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D45250; BAA08207.1; -
DR HSSP; 006323; LAVO.
DR InterPro; IPR003185; PA28_alpha.
DR InterPro; IPR003186; PA28_beta.
DR Pfam; PF02251; PA28_alpha; 1.
DR Pfam; PF02252; PA28_beta; 1.
KW Proteasome; Interferon induction.
SQ SEQUENCE 238 AA; 26857 MW; AFA0F013CECE1CD3 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 238;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 776 EXTLALL 782
Db 97 EXTLALL 103

RESULT 42
PSE2_MOUSE ID_PSE2_MOUSE STANDARD; PRT; 239 AA.

AC P97372; O35562;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Proteasome activator complex subunit 2 (proteasome activator 28-beta subunit) (PA28beta) (PA28b) (Activator of multicatalytic protease subunit 2) (11S regulator complex beta subunit) (REG-beta).
DE subunit 2) (11S regulator complex beta subunit) (REG-beta).
GN (PSME2 OR PA28B1) AND (PSME2B OR PA28B2).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxId=10090;
RN (1)
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6; TISSUE=Spleen;
RC MEDLINE=97364691; PubMed=9218537;
RA Kandil E., Kohda K., Ishibashi T., Tanaka K., Kasahara M.;
RT "PA28 subunits of the mouse proteasome: primary structures and chromosomal localization of the genes.";
RL Immunogenetics 46:337-344(1997).

RN (2)
RP SEQUENCE FROM N.A.
RX STRAIN=B10.BR;
RC MEDLINE=97306277; PubMed=9162094;
RA Jiang H., Monaco J.U.;
RT "Sequence and expression of mouse proteasome activator PA28 and the related autoantigen Ki.";
RL Immunogenetics 46:93-98(1997).
RN (3)
RP SEQUENCE FROM N.A.
RX STRAIN=129/SvJ;
RC MEDLINE=98250122; PubMed=9590240;
RA Kohda K., Ishibashi T., Shibamura N., Tanaka K., Matsuda Y., Kasahara M.;
RT "Characterization of the mouse PA28 activator complex gene family: complete organizations of the three member genes and a physical map of the approximately 150-kb region containing the alpha- and beta-subunit genes.";
RL J. Immunol. 160:4923-4935(1998).

RN (4)
RP SEQUENCE FROM N.A. (PSME2 AND PSME2B).
RX STRAIN=129;
RC MEDLINE=99115499; PubMed=9914329;
RA Li Y., Chambers J., Pang J., Ngo K., Peterson P.A., Leung W.P., Yang Y.;
RT "Characterization of the mouse proteasome regulator PA28b gene.";
RL Immunogenetics 49:149-157(1999).

RN (5)
RP SEQUENCE FROM N.A. (PSME2B).
RX STRAIN=129/SvJ;
RC MEDLINE=99241037; PubMed=10222192;
RA Zais D.M., Klotzel P.M.;
RT "A second gene encoding the mouse proteasome activator PA28beta subunit is part of a LINE1 element and is driven by a LINE1 promoter.";
RL J. Mol. Biol. 287:829-835(1999).

RN (6)
RP SEQUENCE FROM N.A.
RX STRAIN=129/SvJ;
RC MEDLINE=21243724; PubMed=11345588;
RA Yawata M., Murata S., Tanaka K., Ishigatsubo Y., Kasahara M.;
RT "Nucleotide sequence analysis of the ~35-kb segment containing interferon-gamma-inducible mouse proteasome activator genes.";
RL Immunogenetics 53:119-129(2001).

RN (7)
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Embryo;
RC MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Aizawa K., Izawa M., Fukunishi Y., Komno H., Adachi J., Fukuda S., Saio T., Okazaki Y., Gojobori T., Bono H., Kasakawa T., Yamada I., Kadota K., Matsuda H.A., Ashburner W., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Glass C., King B., Kochiwa H., Kuehl P., Lewis S., Matsumoto Y., Nikaide I., Pesole G., Quackenbush J., Schriml L.M., Staahl F., Suzuki R., Tomita M., Wagner L., Washio T.,

RA Sakai K., Okido T., Furuno M., Aono H., Balarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Guetlinich S., Hill D., Hofmann M., Hume D.A., Kamya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Nomberts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seta T., Shihata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsunki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection",
RL Nature 409:685-690(2001).
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mulhally S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences",
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC
CC -1- FUNCTION: Implicated in immunoproteasome assembly and required for
efficient antigen processing. The PA28 activator complex enhances
the generation of class I binding peptides by altering the
cleavage pattern of the proteasome.
CC
CC -1- SUBUNIT: HETERODIMER OF PSME1 AND PSME2, WHICH FORMS A HEXADIMERIC
RING.
CC
CC -1- INDUCTION: By interferon gamma.
CC
CC -1- SIMILARITY: BELONGS TO THE PA28 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; D87910; BAA22040.1; -;
DR EMBL; U60329; AAC53296.1; -;
DR EMBL; AB007138; BAA28837.1; -;
DR EMBL; AF060195; AAC83939.1; -;
DR EMBL; AF155502; AAD38121.1; -;
DR EMBL; AB053120; BAB47405.1; -;
DR EMBL; AK012344; BAB28175.1; -;
DR EMBL; BC005680; AAH05680.1; -;
DR HSSP; Q06323; IAVO.
DR MGD; MGI:1096365; Psme2.
DR MGD; MGI:1341073; Psme2b.
DR InterPro; IPR003185; PA28_alpha.
DR InterPro; IPR003186; PA28_beta.
DR Pfam; PF02251; PA28_alpha; 1.
DR Pfam; PF02252; PA28_beta; 1.
KW Proteasome.
FT CONFLICT 152 152 A -> P (IN REF. 2).
SQ SEQUENCE 239 AA; 27057 MW; D8BB90F78CB0DD7C CRC64;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 776 EXTLALL 782
Db 98 EXTLALL 104
RESULT 43
EL2_RAT
ID EL2_RAT STANDARD; PRT: 271 AA.
AC P00774;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Elastase 2 precursor (EC 3.4.21.71).
GN ELA2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=82182967; PubMed=6918221;
RA McDonald R.J., Swift G.H., Quinto C., Swain W., Picot R.L.,
RA Nikovits W., Rutter W.J.,
RT "Primary structure of two distinct rat pancreatic preproelastases
determined by sequence analysis of the complete cloned messenger
ribonucleic acid sequences",
RL Biochemistry 21:1453-1463(1982).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=85054882; PubMed=6094548;
RA Swift G.H., Craik C.S., Stary S.J., Quinto C., Lahnate R.G.,
RA Rutter W.J., McDonald R.J.,
RT "Structure of the two related elastase genes expressed in the rat
pancreas",
RL J. Biol. Chem. 259:14271-14278(1984).
CC
CC -1- FUNCTION: ACTS UPON ELASTIN.
CC
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Leu-Xaa, Met-Xaa
and Phe-Xaa. Hydrolyzes elastin.
CC
CC -1- SUBCELLULAR LOCATION: Secreted.
CC
CC -1- TISSUE SPECIFICITY: Pancreas.
CC
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. ELASTASE SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; V01233; CAA24543.1; -;
DR EMBL; L00124; AAA98780.1; -;
DR EMBL; L00118; AAA98780.1; JOINED.
DR EMBL; L00119; AAA98780.1; JOINED.
DR EMBL; L00120; AAA98780.1; JOINED.
DR EMBL; L00121; AAA98780.1; JOINED.
DR EMBL; L00122; AAA98780.1; JOINED.
DR EMBL; L00123; AAA98780.1; JOINED.
DR PIR; A00961; ELR2.
DR HSSP; P00772; IELG.
DR MEROPS; S01.155; -;
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; Trypsin; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS02440; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Zymogen; Signal.
FT SIGNAL 1 16
FT PROPEP 17 30 ACTIVATION PEPTIDE.
FT CHAIN 31 271 ELASTASE 2.

Query Match 0.9%; Score 7; DB 1; Length 239;
Best Local Similarity 100.0%; Pred. No. 57;

FT ACT_SITE 75 75 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 123 123 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 218 218 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 60 76 BY SIMILARITY.
 FT DISULFID 157 224 BY SIMILARITY.
 FT DISULFID 188 204 BY SIMILARITY.
 FT DISULFID 214 245 BY SIMILARITY.
 SQ SEQUENCE 271 AA; 28885 MW; 125C783B857B71E3 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 271;
 Best Local Similarity 100.0%; Pred. No. 63;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 219 SNSRTYR 225
 |||||
 Db 78 SNSRTYR 84

RESULT 44
 FPG_OCEIH STANDARD; PRT; 280 AA.
 AC 08EP6;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DE Formamidopyrimidine-DNA glycosylase (EC 3.2.2.23) (Fapy-DNA
 DE glycosylase).
 GN MUTM OR FPG OR OB2162.
 OS Oceanobacillus ihensis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
 OX NCBI_TaxID=182110;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HTE831 / DSM 14371 / JCM 11309;
 RX MEDLINE=22220767; PubMed=12235376;
 RA Takami H., Takaki Y., Uchiyama I.;
 RT "Genome sequence of Oceanobacillus ihensis isolated from the Iheya
 RT Ridges and its unexpected adaptive capabilities to extreme
 RT environments.";
 RL Nucleic Acids Res. 30:3927-3935(2002).
 CC -1- FUNCTION: This enzyme may play a significant role in processes
 CC leading to recovery from mutagenesis and/or cell death by
 CC alkylating agents (By similarity).
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of DNA containing ring-opened N(7)-
 CC methylguanine residues, releasing 2,6-diamino-4-hydroxy-5-(N-
 CC methyl)formamidopyrimidine.
 CC -1- COFACTOR: Binds 1 zinc ion (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE FPG FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, AF004600; BAC14118.1; -.
 DR HAMAP; MF_00103; -; 1.
 DR Pfam; PF01149; Fapy_DNA_glyco; 1.
 DR ProDom; PD003680; Fapy_DNA_glyco; 1.
 DR TIGRPFAM; TIGR00577; fpg; 1.
 DR PROSITE; PS01242; FPG; 1.
 KM DNA repair; Hydrolase; Glycosidase; Zinc; Zinc-finger;
 KW Complete proteome.
 FT ZN FING 249 272 POTENTIAL.
 SQ SEQUENCE 280 AA; 32235 MW; 8E83A26CB786C286 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 280;
 Best Local Similarity 100.0%; Pred. No. 65;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 413 SKLSKOE 419

Db 193 SKLSKOE 199
 |||||

RESULT 45
 LICH_VIBCH STANDARD; PRT; 284 AA.
 ID LICH_VIBCH
 AC 007350;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Lipase chaperone (Lipase foldase) (Lipase helper protein)
 DE (Lipase activator protein) (Lipase modulator).
 GN LIPB OR VCA0222.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 OX NCBI_TaxID=666;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=El Tor O17 / Serotype O1;
 RA Manning P.A.;
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=El Tor N16961 / Serotype O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Ginn M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
 RA McDonald L., Utterback T., Fleischmann R.D., Niemann W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
 RT cholerae.";
 RL Nature 406:477-483(2000).
 CC -1- FUNCTION: SEEMS TO BE ACTING IN THE FOLDING OF THE EXTRACELLULAR
 CC LIPASE DURING ITS PASSAGE THROUGH THE PERIPLASM (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Inner membrane-anchored (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE LIPASE CHAPERONE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, Y00557; CAA68635.1; -.
 DR EMBL; AE004362; AAF96134.1; -.
 DR PIR; D82486; D82486.
 DR TIGR; VCA0222; -.
 DR InterPro; IPR004961; Lipase_chap.
 DR Pfam; PF03280; Lipase_chap; 1.
 KM Lipid degradation; Chaperone; Transmembrane; Periplasmic;
 KW Inner membrane; Complete proteome.
 FT TRANSMEM 4 24 POTENTIAL.
 SQ SEQUENCE 284 AA; 32561 MW; B0064285C85C0BC7 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 284;
 Best Local Similarity 100.0%; Pred. No. 66;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 198 NELSAFE 204
 |||||
 Db 246 NELSAFE 252

RESULT 46
 TPML_BIOGL STANDARD; PRT; 284 AA.
 ID TPML_BIOGL

```

AC 242636;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Tropomyosin 1 (TMI) (BG39).
OS Biomphalaria glabrata (Bloodfluke planorb).
CC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Basommatophora;
CC Lymnaeidae; Planorbidae; Biomphalaria.
CC NCBI_TaxID=6526;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91218806; PubMed=2090946;
RA Dissous C., Torprier G., Duvaux-Miret O., Capron A.;
RT "Structural homology of tropomyosins from the human trematode
RT Schistosoma mansoni and its intermediate host Biomphalaria
RT glabrata."
RL Mol. Biochem. Parasitol. 43:245-255(1990).
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- DOMAIN: THE MOLECULE IS IN A COILED COIL STRUCTURE. THE SEQUENCE
CC EXHIBITS A PROMINENT SEVEN-RESIDUES PERIODICITY.
CC -1- SIMILARITY: BELONGS TO THE TROPOMYOSIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M85199; AAA27817.1; -.
DR PIR; A33085; A33085.
DR InterPro; IPR002017; Spectrin.
DR InterPro; IPR000533; Tropomyosin.
DR Pfam; PF00261; Tropomyosin; 1.
DR PRINTS; PR00194; TROPOMYOSIN.
DR Coiled coil; Repeat; Multigene family.
DR Coiled coil; Repeat; Multigene family.
SQ SEQUENCE 284 AA; 32716 MW; 13735417CE49A8 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 284;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 728 EAOLKEA 734
DB 145 EAOLKEA 151

RESULT 47
TPM2_BIOGL STANDARD; PRT; 284 AA.
AC P43689;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Tropomyosin 2 (TMI2).
OS Biomphalaria glabrata (Bloodfluke planorb).
CC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Basommatophora;
CC Lymnaeidae; Planorbidae; Biomphalaria.
CC NCBI_TaxID=6526;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Brazilian;
RA MEDLINE=93292622; PubMed=7685709;
RA Weston D.S., Kemp W.M.;
RT "Schistosoma mansoni: comparison of cloned tropomyosin antigens
RT shared between adult parasites and Biomphalaria glabrata."
RL Exp. Parasitol. 76:358-370(1993).
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- DOMAIN: THE MOLECULE IS IN A COILED COIL STRUCTURE. THE SEQUENCE
CC EXHIBITS A PROMINENT SEVEN-RESIDUES PERIODICITY.
CC -1- SIMILARITY: BELONGS TO THE TROPOMYOSIN FAMILY.
-----

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M97554; AAA27816.1; -.
DR InterPro; IPR002017; Spectrin.
DR InterPro; IPR000533; Tropomyosin.
DR Pfam; PF00261; Tropomyosin; 1.
DR PRINTS; PR00194; TROPOMYOSIN.
DR Coiled coil; Repeat; Multigene family.
DR Coiled coil; Repeat; Multigene family.
SQ SEQUENCE 284 AA; 32683 MW; 256103712D9B4C33 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 284;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 728 EAOLKEA 734
DB 145 EAOLKEA 151

RESULT 48
TPM2_TRICO STANDARD; PRT; 284 AA.
AC P15846;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Tropomyosin, muscle.
OS Trichostrongylus colubriformis (Black scour worm).
CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdida; Strongylida;
CC Trichostrongylidae; Trichostrongylinae;
CC Trichostrongylus.
CC NCBI_TaxID=6319;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90114326; PubMed=2514356;
RX Frenkel M.J., Savin K.W., Bakker R.E., Ward C.W.;
RA "Characterization of cDNA clones coding for muscle tropomyosin of the
RT nematode Trichostrongylus colubriformis."
RL Mol. Biochem. Parasitol. 37:191-200(1989).
RN [2]
RP PARTIAL SEQUENCE.
RX MEDLINE=89339870; PubMed=2759773;
RX O'Donnell I.J., Dineen J.K., Wagland B., Letho S., Werkmeister J.A.,
RA Ward C.W.;
RT "A novel host-protective antigen from Trichostrongylus
RT colubriformis."
RL Int. J. Parasitol. 19:327-335(1989).
CC -1- FUNCTION: TROPOMYOSIN, IN ASSOCIATION WITH THE TROPONIN COMPLEX,
CC PLAYS A CENTRAL ROLE IN THE CALCIUM DEPENDENT REGULATION OF
CC STRIATED MUSCLE CONTRACTION.
CC -1- SUBUNIT: Homodimer.
CC -1- DEVELOPMENTAL STAGE: PRESENT IN L3 (THIRD STAGE), L4 AND ADULT
CC WORMS.
CC -1- DOMAIN: THE MOLECULE IS IN A COILED COIL STRUCTURE. THE SEQUENCE
CC EXHIBITS A PROMINENT SEVEN-RESIDUES PERIODICITY.
CC -1- SIMILARITY: BELONGS TO THE TROPOMYOSIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; J04669; AAA30103.1; -.
DR InterPro; IPR000533; Tropomyosin.
-----

```

DR Pfam: PF00261; Tropomyosin; 1.
 DR PRINTS; PR00194; TROPOMYOSIN.
 DR PROSITE; PS00326; TROPOMYOSIN; 1.
 KW Muscle protein; Coiled coil; Repeat.
 FT DOMAIN 1 284 T -> R (IN REF. 2).
 FT CONFLICT 277 277 COILED COIL.
 SQ SEQUENCE 284 AA; 33050 MW; 230573BEE2449C1 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 284;
 Best Local Similarity 100.0%; Pred. No. 66;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 728 EAQKKEA 734
 |||||
 Db 145 EAQKKEA 151

RESULT 49
 ID LECT1 CLALU STANDARD; PRT; 293 AA.
 AC 039528;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Agglutinin I precursor (ClAI) (LeccIAI).
 OS Cladastis lutea (Yellow wood).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eurosid 1; Finales; Fabaceae; Papilionoideae; Sophoreae; Cladastis.
 NCBI_TaxID=38412;
 RX MEDLINE=6123235; PubMed=8534854;
 RP TISSUE=Barik;
 RA van Damme E.J.M., Barre A., Bemer V., Rouge P., van Leuven F.,
 RA Peumans W.J.;
 RA "A lectin and a lectin-related protein are the two most prominent
 RT proteins in the bark of yellow wood (Cladastis lutea).";
 RL Plant Mol. Biol. 29:579-598(1995).

CC -1- FUNCTION: NANNOSE/GLUCOSE BINDING BARK LECTIN.
 CC -1- FUNCTION: BARK LECTINS ARE STORAGE PROTEIN THAT PROBABLY MAINTAINS
 CC STOCKS OF NITROGEN DURING DORMANT PERIOD. SELF-AGGREGATABLE
 CC MOLECULES THAT CAN BIND THEIR OWN CARBOHYDRATE SIDE CHAINS. THEY
 CC COULD ALSO PLAY A ROLE IN THE PLANT'S DEFENCE AGAINST PHYTOPHAGOUS
 CC INVERTEBRATES OR HERBIVOROUS HIGHER ANIMALS.
 CC -1- SUBUNIT: HOMOTETRAMER OF FOUR 32 KDa MONOMERS WHICH ARE POST-
 CC TRANSLATIONALLY CLEAVED INTO A TWO SUBUNITS: A AND B.
 CC -1- MISCELLANEOUS: BINDS ONE MANGANESE (OR OTHER TRANSITION METAL) ION
 CC AND ONE CALCIUM ION. THE METAL IONS ARE ESSENTIAL FOR THE
 CC SACCHARIDE-BINDING AND CELL-AGGLUTININATING ACTIVITIES (BY
 CC SIMILARITY).
 CC -1- SIMILARITY: TO OTHER LEGUMINOUS LECTINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; U21958; AAC49136.1; -
 DR PIR; S66356; S66356.
 DR HSSP; P19588; ILU.
 DR InterPro; IPR000985; Lectin_lega.
 DR InterPro; IPR001220; Lectin_legb.
 DR Pfam; PF00138; lectin_lega; 1.
 DR Pfam; PF00139; lectin_legb; 1.
 DR ProDom; PD000671; Lectin_lega; 1.
 DR ProDom; PD000711; Lectin_legb; 1.
 DR PROSITE; PS00307; LECTIN_LEGME_BETA; 1.
 DR PROSITE; PS00308; LECTIN_LEGME_ALPHA; 1.
 KW Lectin; Calcium; Manganese; Glycoprotein; Signal; Mannose-binding.

FT SIGNAL 1 35
 FT CHAIN 36 161
 FT CHAIN 162 293
 FT CHAIN 171 171
 FT METAL 173 173
 FT METAL 175 175
 FT METAL 177 177
 FT METAL 181 181
 FT METAL 185 186
 FT CARBOHYD 64 64
 FT CARBOHYD 152 152
 FT CARBOHYD 162 162
 FT CONFLICT 47 47
 FT CONFLICT 162 162
 SQ SEQUENCE 293 AA; 32128 MW; EDBBD3FF5FA3C6C CRC64;

Query Match 0.9%; Score 7; DB 1; Length 293;
 Best Local Similarity 100.0%; Pred. No. 68;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 424 TLTAKE 430
 |||||
 Db 279 TLTAKE 285

RESULT 50
 ID TAL_BUCAP STANDARD; PRT; 316 AA.
 AC 08KA27;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Transaldolase (EC 2.2.1.2).
 GN TAL OR BUSG085.
 OS Buchnera aphidicola (subsp. Schizaphis graminum).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Buchnera.
 NCBI_TaxID=98794;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22084549; PubMed=12089438;
 RA Tamás I., Klasson L., Canhaeck B., Naeslund A.K., Eriksson A.-S.,
 RA Werngren J.J., Sandström J.P., Moran N.A., Anderson S.G.E.;
 RT "50 million years of genomic stasis in endosymbiotic bacteria.";
 RL Science 296:2376-2379(2002).

CC -1- FUNCTION: Transaldolase is important for the balance of
 CC metabolites in the pentose-phosphate pathway (By similarity).
 CC -1- CATALYTIC ACTIVITY: Sedoheptulose 7-phosphate + D-glyceraldehyde
 CC 3-phosphate = D-erythrose 4-phosphate + D-fructose 6-phosphate.
 CC -1- PATHWAY: Pentose phosphate pathway; nonoxidative part.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -1- SIMILARITY: BELONGS TO THE TRANSALDOLASE FAMILY. SUBFAMILY 1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AE014084; AM67655.1; -
 DR HAMAP; MF_00492; - 1.
 DR InterPro; IPR001585; Transaldolase.
 DR InterPro; IPR004730; Transaldolase_AB.
 DR Pfam; PF00923; Transaldolase; 1.
 DR TIGRfam; TIGR00874; talAB; 1.
 DR PROSITE; PS01054; TRANSALDOLASE_1; 1.
 DR PROSITE; PS00958; TRANSALDOLASE_2; FALSE_NEG.
 KW Transferrase; Pentose shunt; Complete proteome.
 FT ACT SITE 131 131
 FT ACT SITE 131 131
 SQ SEQUENCE 316 AA; 35825 MW; 2BF7A4E4CC3425B0 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 316;
 Best Local Similarity 100.0%; Pred. No. 73;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 390 EKGISR 396
 |||||
 Db 119 EKGISR 125

RESULT 51

PTPA_RABIT STANDARD; PRT; 323 AA.
 ID PTPA_RABIT
 AC Q28717;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Protein phosphatase 2A, regulatory subunit B' (PP2A, subunit B', PRS3 isoform) (phosphotyrosyl phosphatase activator) (PTPA).
 GN PPP2R4 OR PTPA.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Skeleral muscle;
 RX MEDLINE=94253154; PubMed=8195217;
 RA Cayla X., Van Hoof C., Bosch M., Waalkens E., Peeters B., Meilevede W., Goris J.;
 RT "Molecular cloning, expression, and characterization of PTPA, a protein that activates the tyrosyl phosphatase activity of protein phosphatase 2A.";
 RT J. Biol. Chem. 269:15668-15675(1994).
 RU
 CC - FUNCTION: REVERSIBLY STIMULATES THE VARIABLE PHOSPHOTYRSL PHOSPHATASE ACTIVITY OF PP2A CORE HETERODIMER IN PRESENCE OF ATP AND MG(2+) (IN VITRO).
 CC - SUBUNIT: ASSOCIATES WITH PP2A HETERODIMERIC CORE ENZYME, COMPOSED OF A 36 KDA CATALYTIC SUBUNIT (SUBUNIT C) AND A 65 KDA CONSTANT REGULATORY SUBUNIT (PP2S OR SUBUNIT A).
 CC - TISSUE SPECIFICITY: Widely expressed.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X73479; CAA51874.1; -.
 DR PIR: B54021; B54021.
 DR InterPro: IPR004327; PhosTy_phatase_ac.
 DR Pfam: PF03095; PTPA.1.
 DR SEQUENCE 323 AA; 36587 MW; 808F404934F3D85 CRC64;
 SQ
 Query Match 0.9%; Score 7; DB 1; Length 323;
 Best Local Similarity 100.0%; Pred. No. 74;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 776 EKLALL 782
 |||||
 Db 75 EKLALL 81

OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;

RP SEQUENCE FROM N.A.
 RX MEDLINE=95254648; PubMed=7736590;
 RA Preker P.J., Lingner J., Manivelle-Sebastia L., Keller W.;
 RT "The FIP1 gene encodes a component of a yeast pre-mRNA polyadenylation factor that directly interacts with poly(A) polymerase.";
 RT Cell 81:379-389(1995).
 RL Cell 81:379-389(1995).
 RN [2]

RP SEQUENCE FROM N.A.
 RA Ramezani Rad M., Kirchbach L., Hollenberg C.P.;
 RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
 CC - FUNCTION: PRE-MRNA POLYADENYLATION FACTOR THAT DIRECTLY INTERACTS WITH POLY(A) POLYMERASE. SEEMS TO TETHER PAPI TO THE CLEAVAGE FACTOR I. INTERACTS WITH RNA14.
 CC - SUBCELLULAR LOCATION: Nuclear.
 CC - SIMILARITY: TO S.POMBE SPAC2267.10.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X83796; CAA58727.1; -.
 DR EMBL: Z49593; CAA89621.1; -.
 DR PIR: A56545; A56545.
 DR SGD: S0003853; FIP1.
 DR GO: GO:0006379; P:RNA cleavage, IDA.
 DR GO: GO:0006378; P:RNA polyadenylation, IDA.
 DR Pfam: PF05182; Fip1.1.
 KW mRNA processing; Nuclear protein.
 FT DOMAIN 61 80 ASP-RICH (ACIDIC).
 FT DOMAIN 284 318 PRO-RICH.
 FT DOMAIN 258 266 POLY-ASN.
 FT DOMAIN
 SQ SEQUENCE 327 AA; 35777 MW; 471CA2B0CDF99D0A CRC64;
 Query Match 0.9%; Score 7; DB 1; Length 327;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 709 PVEETPA 715
 |||||
 Db 51 PVEETPA 57

RESULT 53

KAPR_DICD1 STANDARD; PRT; 327 AA.
 ID KAPR_DICD1
 AC P05987;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-NOV-1988 (Rel. 09, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE CAMP-dependent protein kinase regulatory chain.
 GN PKAR.
 OS Dictyostelium discoideum (Slime mold).
 CC Eukaryota; Mycetozoa; Dictyostelidae; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87092396; PubMed=3467359;
 RA Metzger R., Lacombe M.-E., Simon M.-N., de Gunzburg J., Veron M.;
 RT "Cloning and cDNA sequence of the regulatory subunit of CAMP-dependent protein kinase from Dictyostelium discoideum.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:6-10(1987).
 RN [2]
 RP MUTANT RDEC.

```

RX MEDLINE-92186968; PubMed1312226;
RA Simon M.-N., Pellegrini O., Veron M., Kay R.R.;
RT "Mutation of protein kinase A causes heterochronic development of
RL Dicyostelium.";
RL Nature 356:171-172 (1992).
CC -1- SUBUNIT: IN DICTYOSTELIUM THE HOLOENZYME IS A DIMER COMPOSED OF
CC A REGULATORY (R) AND A CATALYTIC (C) SUBUNIT. IN THE PRESENCE OF
CC CAMP IT DISSOCIATES INTO THE ACTIVE C SUBUNIT AND AN R MONOMER.
CC (IN OTHER EUKARYOTES THE HOLOENZYME IS A TETRAMER COMPOSED OF 2
CC REGULATORY (R) & 2 CATALYTIC (C) SUBUNITS. IN THE PRESENCE OF CAMP
CC IT DISSOCIATES INTO ACTIVE MONOMERIC C SUBUNITS AND AN R DIMER).
CC -1- DOMAIN: LACKS THE N-TERMINAL DOMAIN REQUIRED FOR THE ASSOCIATION
CC OF REGULATORY SUBUNIT INTO DIMERS IN OTHER EUKARYOTES.
CC -1- PTM: THE PSEUDOPHOSPHORYLATION SITE BINDS TO THE SUBSTRATE-
CC BINDING REGION OF THE CATALYTIC CHAIN BUT IS NOT PHOSPHORYLATED.
CC THE PHYSIOLOGICAL SIGNIFICANCE OF PHOSPHORYLATIONS BY OTHER
CC KINASES IS UNCLEAR.
CC -1- MISCELLANEOUS: IN DICTYOSTELIUM EACH R SUBUNIT CARRIES ONLY 1
CC HIGH-AFFINITY CAMP BINDING SITE (2 IN OTHER EUKARYOTES).
CC -1- MISCELLANEOUS: IN RBC MUTANTS AGGREGATION IS FOLLOWED ABRUPTLY BY
CC THE MATURATION OF SPORE AND STALK CELLS. THIS IS A DUE TO
CC MUTATION(S) IN THE R SUBUNIT. THE MUTANT R SUBUNIT BINDS CAMP BUT
CC INHIBITS POORLY THE ENZYMATIC ACTIVITY OF THE C SUBUNIT.
CC -1- SIMILARITY: CONTAINS 2 CYCLIC NUCLEOTIDE-BINDING DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE CAMP-DEPENDENT KINASE REGULATORY CHAIN
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M15081; AAA3236.1; -
CC PIR: A29076; OKDRC.
CC DR HSSP: P00514; IAPK.
CC DR DICTYDB: D02006; PKAR.
CC DR InterPro: IPR002373; CAMP_kin.
CC DR InterPro: IPR000595; CNMP_binding.
CC DR Pfam: PF00027; CNMP_binding; 2.
CC DR PRINTS: PR00103; CAMPKINASE.
CC DR SMART: SMO0100; CNMP; 2.
CC DR PROSITE: PS00888; CNMP_BINDING_1; 2.
CC DR PROSITE: PS00889; CNMP_BINDING_2; 2.
CC DR PROSITE: PS00942; CNMP_BINDING_3; 2.
CC DR CAMP-binding; Phosphorylation; Repeat.
CC KW DOMAIN 1 65 DIMERIZATION AND PHOSPHORYLATION.
CC FT SITE 27 31 PSEUDOPHOSPHORYLATION MOTIF.
CC FT NP_BIND 66 188 CAMP (HIGH AFFINITY).
CC FT NP_BIND 189 327 CAMP (LOW AFFINITY).
CC FT REPEAT 66 188 1.
CC FT REPEAT 189 327 2.
CC FT MOD_RES 32 32 PHOSPHORYLATION (BY SIMILARITY).
CC FT BINDING 136 136 CAMP.
CC FT BINDING 145 145 CAMP.
CC FT BINDING 262 262 CAMP.
CC FT BINDING 271 271 CAMP.
CC FT VARIANT 30 30 A -> T (IN RBC MUTANT).
CC FT VARIANT 327 AA; 36836 MW; D76BF40E3FE22CA1 CRC64;
CC SQ SEQUENCE

```

Query Match 0.9%; Score 7; DB 1; Length 327;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 724 TEKVEAQ 730
 |||||
 DB 12 TEKVEAQ 18

RESULT 54
 G3P2_AGABI

```

ID G3P2_AGABI STANDARD; PRT; 338 AA.
AC P32636;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glyceraldehyde 3-phosphate dehydrogenase 2 (EC 1.2.1.12) (GAPDH 2).
GN GPD2.
OS Agaricus bisporus (Common mushroom).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Agaricaceae; Agaricus.
OX NCBI_Taxid=5341;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Hors U3;
RX MEDLINE-93113715; PubMed1473176;
RA Harnsen M.C., Schuren F.H.J., Mousha S.M., van Zuijen C.M., Punt P.J.,
RA Wessels J.G.H.;
RT "Sequence analysis of the glyceraldehyde-3-phosphate dehydrogenase
RT genes from the basidiomycetes Schizophyllum commune, Phanerochaete
RT chrysosporium and Agaricus bisporus.";
RL Curr. Genet. 22:447-454(1992).
CC -1- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate + phosphate +
CC NAD(+) = 3-phospho-D-glyceroyl phosphate + NADH.
CC -1- PATHWAY: Second phase of glycolysis; first step.
CC -1- SUBUNIT: Homotetramer.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the glyceraldehyde 3-phosphate
CC dehydrogenase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M81728; AAA32634.1; -
CC PIR: S26976; S26976.
CC DR HSSP: P00357; 4GPD.
CC DR InterPro: IPR000173; GAP_dhndrogenase.
CC DR InterPro: IPR006424; GAPDH-I.
CC DR Pfam: PF00044; gpdh; 1.
CC DR Pfam: PF02800; gpdh; C; 1.
CC DR PRINTS: PR00078; G3PDHRCGNASE.
CC DR TIGRPFAMs: TIGR01534; GAPDH-1; 1.
CC DR PROSITE: PS00071; GAPDH; 1.
CC KW Glycolysis; Oxidoreductase; NAD; Multigene family.
CC FT BINDING 150 150 GLYCERALDEHYDE 3-PHOSPHATE.
CC FT ACT_SITE 177 177 ACTIVATES THIOL GROUP DURING CATALYSIS.
CC SQ SEQUENCE 338 AA; 36570 MW; 3381DBBF73B84469 CRC64;

```

Query Match 0.9%; Score 7; DB 1; Length 338;
 Best Local Similarity 100.0%; Pred. No. 77;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 621 TVEVNGK 627
 |||||
 DB 57 TVEVNGK 63

RESULT 55
 G3P_SCHMA
 ID G3P_SCHMA STANDARD; PRT; 338 AA.
 AC P20287;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Glyceraldehyde 3-phosphate dehydrogenase (EC 1.2.1.12) (GAPDH) (Major
 DE larval surface antigen) (P-37).
 OS Schistosoma mansoni (Blood fluke).
 OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
 OC Schistosomatidae; Schistosomatidae; Schistosoma.

```

OX NCBI_TaxID=6183;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Puerto Rican;
RX MEDLINE=90063470; PubMed=2584935;
RA Gondot-Crozal V., Caillol D., Djabali M., Dessein A.J.;
RT "The major parasite surface antigen associated with human resistance
  to schistosomiasis is a 37-kD glyceraldehyde-3p-dehydrogenase.";
RU J. Exp. Med. 170:2065-2080(1989)
CC - FUNCTION: THIS ANTIGEN IS ASSOCIATED WITH HUMAN RESISTANCE TO
  SCHISTOSOMIASIS.
CC - CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate + phosphate +
  NAD(+) = 3-phospho-D-glyceroyl phosphate + NADH.
CC - PATHWAY: Second phase of glycolysis; first step.
CC - SIMILARITY: Belongs to the glyceraldehyde 3-phosphate
  dehydrogenase family.
CC PIR: JLO121; JLO121.
DR HSSP; P56649; 1DSS.
DR InterPro; IPR000173; GAP_dhrogenase.
DR InterPro; IPR006424; GAPDH-1.
DR Pfam; PF00044; gpdh; 1.
DR Pfam; PF02800; gpdh; 1.
DR PRINTS; PR00078; G3PDHGRNASE.
DR TIGRFAMs; TIGR01534; GAPDH-1; 1.
DR PROSITE; PS00071; GAPDH; 1.
KW Glycolysis; Oxidoreductase; NAD; Antigen.
FT BINDING 153 153
FT ACT_SITE 180 180
FT FT SIMILARITY.
FT ACTIVATES THIOL GROUP DURING CATALYSIS
  (BY SIMILARITY).
SQ SEQUENCE 338 AA; 36377 MW; 3DAC70E93C1E24BE CRC64;

Query Match 0.9%; Score 7; DB 1; Length 338;
Best local Similarity 100.0%; Pred. No. 77;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 509 EYTEDV 515
DB 276 EYTEDV 282

```

```

RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Oucello B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmlel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Garson J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
  Agrobacterium tumefaciens C58.";
RU Science 294:2323-2328(2001).
CC - CATALYTIC ACTIVITY: Uroporphyrinogen-III = coproporphyrinogen + 4
  CO(2).
CC - PATHWAY: Porphyrin biosynthesis.
CC - SUBCELLULAR LOCATION: Cytoplasmic (probable).
CC - SIMILARITY: BELONGS TO THE UROPORPHYRINOGEN DECARBOXYLASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.ebi.ac.uk/announcements
  or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; AE009230; AAL43816.1; -
CC EMBL; AE008196; AAK8546.1; -
CC PIR; A97699; A97699.
CC PIR; AB2925; AB2925.
DR HAMAP; MF_00218; -. 1.
DR InterPro; IPR006361; Heme.
DR InterPro; IPR000257; Uro-decarboxyls.
DR Pfam; PF01208; URO-D; 1.
DR ProDom; PD003225; Uro-decarboxyls; 1.
DR TIGRFAMs; TIGR01464; Heme; 1.
DR PROSITE; PS00906; UROD_1; 1.
DR PROSITE; PS00907; UROD_2; 1.
DR Lyase; Decarboxylase; Porphyrin biosynthesis; Complete proteome.
KW SEQUENCE 344 AA; 37173 MW; 96351DE3393ECB46 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 344;
Best local Similarity 100.0%; Pred. No. 78;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 776 EKLALL 782
DB 174 EKLALL 180

```

CC TISSUE=Blood;
 CC MEDLINE=96015056; PubMed=8530035;
 CC Van Hoof C., Aly M., Garcia A., Cayla X., Caesman J., Merlevede W.,
 CC Goris J.;
 CC "Structure and chromosomal localization of the human gene of the
 CC phosphotyrosyl phosphatase activator (PTPA) of protein phosphatase
 CC 2A.";
 CC Genomics 28:261-272(1995).
 CC [3]
 CC SEQUENCE FROM N.A. (ISOFORMS 2; 3 AND 4).
 CC MEDLINE=20341117; PubMed=10880954;
 CC Janssens V., van Hoof C., Martens E., de Baere I., Merlevede W.,
 CC Goris J.;
 CC "Identification and characterization of alternative splice products
 CC encoded by the human phosphotyrosyl phosphatase activator gene.";
 CC Eur. J. Biochem. 267:4406-4413(2000).
 CC [4]
 CC SEQUENCE FROM N.A. (ISOFORM 1).
 CC TISSUE=Placenta;
 CC MEDLINE=22388257; PubMed=12477932;
 CC Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,
 CC Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Bhat N.K.,
 CC Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 CC Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Hsieh F.,
 CC Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 CC Stapleton M., Soares M.B., Bonaldo M.F., Casavant P., Prange C.T.,
 CC Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Mullan S.J.,
 CC Raha S.S., Loughellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 CC Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 CC Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 CC Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 CC Fahy J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,
 CC Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,
 CC Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 CC Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,
 CC Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smalios D.E.,
 CC Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 CC "Generation and initial analysis of more than 15,000 full-length
 CC human and mouse cDNA sequences.";
 CC Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -I- FUNCTION: REVERSIBLY STIMULATES THE VARIABLE PHOSPHOTYROSYL
 CC PHOSPHATASE ACTIVITY OF PP2A CORE HETERODIMER IN PRESENCE OF ATP
 CC AND MG(2+) (IN VITRO).
 CC -I- SUBUNIT: ASSOCIATES WITH PP2A HETERODIMERIC CORE ENZYME, COMPOSED
 CC OF A 36 KDA CATALYTIC SUBUNIT (SUBUNIT C) AND A 65 KDA CONSTANT
 CC REGULATORY SUBUNIT (PR65 OR SUBUNIT A).
 CC -I- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=4;
 CC Name=2; Synonyms=beta;
 CC IsoId=Q15257-1; Sequence=Displayed;
 CC Name=1; Synonyms=Alpha;
 CC IsoId=Q15257-2; Sequence=VSP_005123;
 CC Name=3; Synonyms=Delta;
 CC IsoId=Q15257-3; Sequence=VSP_005122;
 CC Name=4; Synonyms=Epsilon;
 CC IsoId=Q15257-4; Sequence=VSP_005124;
 CC -I- TISSUE SPECIFICITY: Widely expressed.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; X86436; CAA60163.1; JOINED.
 DR EMBL; X86437; CAA60163.1; JOINED.
 DR EMBL; X86438; CAA60163.1; JOINED.
 DR EMBL; X86439; CAA60163.1; JOINED.
 DR EMBL; X86428; CAB77601.1; JOINED.
 DR EMBL; X86429; CAB77601.1; JOINED.
 DR EMBL; X86430; CAB77601.1; JOINED.
 DR EMBL; X86431; CAB77601.1; JOINED.
 DR EMBL; X86432; CAB77601.1; JOINED.
 DR EMBL; X86433; CAB77601.1; JOINED.
 DR EMBL; X86434; CAB77601.1; JOINED.
 DR EMBL; X86435; CAB77601.1; JOINED.
 DR EMBL; X86436; CAB77601.1; JOINED.
 DR EMBL; X86437; CAB77601.1; JOINED.
 DR EMBL; X86438; CAB77601.1; JOINED.
 DR EMBL; X86439; CAB77601.1; JOINED.
 DR EMBL; X86428; CAB77602.1; JOINED.
 DR EMBL; X86429; CAB77602.1; JOINED.
 DR EMBL; X86432; CAB77602.1; JOINED.
 DR EMBL; X86434; CAB77602.1; JOINED.
 DR EMBL; X86435; CAB77602.1; JOINED.
 DR EMBL; X86436; CAB77602.1; JOINED.
 DR EMBL; X86437; CAB77602.1; JOINED.
 DR EMBL; X86438; CAB77602.1; JOINED.
 DR EMBL; X86439; CAB77602.1; JOINED.
 DR EMBL; X86428; CAB77603.1; JOINED.
 DR EMBL; X86429; CAB77603.1; JOINED.
 DR EMBL; X86430; CAB77603.1; JOINED.
 DR EMBL; X86434; CAB77603.1; JOINED.
 DR EMBL; X86435; CAB77603.1; JOINED.
 DR EMBL; X86436; CAB77603.1; JOINED.
 DR EMBL; X86437; CAB77603.1; JOINED.
 DR EMBL; X86438; CAB77603.1; JOINED.
 DR EMBL; X86439; CAB77603.1; JOINED.
 DR EMBL; X86428; CAB77603.1; JOINED.
 DR EMBL; BC002545; AAH02545.1; -.
 DR EMBL; BC011605; AAH11605.1; -.
 DR PIR; A54021; A54021.
 DR GeneW; HGNC:9308; PPR2R4.
 DR MIM; 600756; -.
 DR GO; GO:0005625; C:soluble fraction; TAS.
 DR GO; GO:0008601; P:protein phosphatase type 2A, regulator acti. .; NAS.
 DR GO; GO:0008160; P:protein tyrosine phosphatase activator acti. .; TAS.
 DR GO; GO:0006470; P:protein amino acid dephosphorylation; NAS.
 DR InterPro; IPR004327; Phstyr_phaste_ac.
 DR Pfam; PF03095; PTPA; 1.
 KW Alternative splicing
 FT VARSPPLIC 45 108 Missing (in isoform 3).
 FT VARSPPLIC 73 107 /FTid=VSP_005122.
 FT VARSPPLIC 73 107 Missing (in isoform 1).
 FT VARSPPLIC 73 149 /FTid=VSP_005123.
 FT VARSPPLIC 73 149 Missing (in isoform 4).
 FT CONFLICT 113 113 /FTid=VSP_005124.
 FT CONFLICT 297 297 L->V (IN REF. 4).
 FT CONFLICT 357 357 MISSING (IN REF. 2 AND 3).
 SQ SEQUENCE 358 AA; 40681 MW; 6A99C521AF5F6BB CRC64;
 Query Match Score 7; DB 1; Length 358;
 Best Local Similarity 100.0%; Pred. No. 81;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 776 EXTLALL 782
 Db 110 EXTLALL 116
 RESULT 58
 PHEA.ECOLI STANDARD; PRT; 386 AA.
 AC P07022; P78204;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE P-protein [Includes: Chorismate mutase (EC 5.4.99.5) (CM); Prephenate

```

DE dehydratase (EC 4.2.1.51) (PDT)].
GN PHEA OR B2599 OR Z3891 OR EC53462 OR SF2659.
OS Escherichia coli.
OC Escherichia coli O157:H7, and
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334, 623;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=K12;
RX MEDLINE=85134883; PubMed=6396419;
RA Hudson G.S., Davidson B.E.;
RT "Nucleotide sequence and transcription of the phenylalanine and
RL tyrosine operons of Escherichia coli K12.";
RN J. Mol. Biol. 180:1023-1051(1984).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor M., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,
RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubraman S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT -K12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features.";
RL DNA Res. 4:91-113(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Pocanousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [5]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsuka E., Nakayama K., Morita T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kunita S., Shiba T., Hattori M., Shingawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [6]
RP SEQUENCE OF 1-50 FROM N.A.
RC SPECIES=E.coli;
RX MEDLINE=91072346; PubMed=2254312;
RA Gavini N., Davidson B.E.;
RT "pneao mutants of Escherichia coli have a defective pheA attenuator.";
RN J. Biol. Chem. 265:21532-21535(1990).
RN [7]
RP SEQUENCE OF 1-9 FROM N.A.
RC SPECIES=E.coli;
RX MEDLINE=79033820; PubMed=3602214;
RA Zurawski G.R., Brown K., Killingly D., Yanofsky C.;
RT "Nucleotide sequence of the leader region of the phenylalanine operon
RT of Escherichia coli";
RL Proc. Natl. Acad. Sci. U.S.A. 75:4271-4275(1978).
RN [8]
RP SEQUENCE FROM N.A.
RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xie Y., Zhao A., Gao Y., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441(2002).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1-109.
RN Nucleic Acids Res. 30:4432-4441(2002).
RN [10]
RP SPECIES=E.coli;
RX Lee A.Y., Karplus P.A., Ganem B., Clardy J.;
RT "Atomic structure of the buried catalytic pocket of Escherichia coli
RT chorismate mutase.";
RL J. Am. Chem. Soc. 117:3627-3628(1995).
RN [10]
RP DOMAIN.
RC SPECIES=E.coli;
RX MEDLINE=96165805; PubMed=9497350;
RA Zhang S., Pohnert G., Kongaer P., Wilson D.B., Clardy J.,
RA Ganem B.;
RT "Chorismate mutase-prephenate dehydratase from Escherichia coli: Study
RT of catalytic and regulatory domains using genetically engineered
RT proteins";
RL J. Biol. Chem. 273:6248-6253(1998).
RN [11]
RP -1- CATALYTIC ACTIVITY: Chorismate = prephenate.
CC -1- CATALYTIC ACTIVITY: Prephenate = phenylpyruvate + H(2)O + CO(2).
CC -1- PATHWAY: L-phenylalanine biosynthesis.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: TO OTHER PDT ENZYMES OR DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M10431; AAA2430.1; -
DR EMBL: AE000346; AAC75648.1; -
DR EMBL: D90887; BAA16482.1; -
DR EMBL: D90888; BAA16484.1; -
DR EMBL: AE005490; AAG57710.1; -
DR EMBL: AP002562; BAB36885.1; -
DR EMBL: M58024; AAA62784.1; -
DR EMBL: V00314; CAA23601.1; -
DR EMBL: AE015281; AAN44155.1; -
DR PIR: A30261; KMECPW.
DR PIR: F91061; F91061.
DR PDB: 1ECM; 01-DEC-95.
DR EcoGene: EG10707; pheA.
DR InterPro: IPR002701; Chorismate mut.
DR InterPro: IPR001086; Preph. dehydratase.
DR Pfam: PF01817; Chorismate_mut; 1.
DR Pfam: PF00800; PDT; 1.
DR PIRSF: PIRSF001500; Chor mut. pdt. Ppr; 1.
DR PROSITE: PS00857; PREPHENATE DEHYDR 1; 1.
DR PROSITE: PS00858; PREPHENATE DEHYDR 2; 1.
KM Phenylalanine biosynthesis; Isomerase; Lyase; Multifunctional enzyme;
KM 3D-structure; Complete proteome.
FT DOMAIN 1 109 CHORISMATE MUTASE.
FT DOMAIN 110 285 PREPHENATE DEHYDRATASE.
FT DOMAIN 286 386 REGULATORY (PHE-BINDING).

```



```

FT ACT_SITE 278 278 ESSENTIAL FOR PREPHENATE DEHYDRATASE
FT HELIX 7 42 ACTIVITY (POTENTIAL).
FT TURN 43 43
FT HELIX 49 65
FT TURN 66 67
FT HELIX 70 99
SQ SEQUENCE 386 AA; 43111 MW; 4B0960854C75A4F1 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 386;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 776 EKLALL 782
DB 19 EKLALL 25

RESULT 59
YHFX_ECOLI STANDARD; PRT; 387 AA.
AC P45550:
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein yHfx.
GN YHFX OR B3381.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxId=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R.; Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; U18997; AAA58178.1; -
CC DR EMBL; AE000413; AAC76406.1; -
CC DR PIR; H65132; H65132.
CC DR Ecocore; EG12919; yHfx.
CC DR Pfam; PF01168; Ala racemase N; 1.
CC KW Hypothetical protein; Complete proteome.
CC SQ SEQUENCE 387 AA; 42294 MW; EEF7881339ED5971 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 387;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 775 AEKLAL 781
DB 70 AEKLAL 76

RESULT 60
PGK_THETN STANDARD; PRT; 394 AA.
AC Q8R965;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)

```

```

DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phosphoglycerate kinase (EC 2.7.2.3).
GN PGK OR TTE1761.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxId=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4 / JCM 11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
CC -1- CATALYTIC ACTIVITY: ATP + 3-phospho-D-glycerate = ADP + 3-
CC phospho-D-glyceroyl phosphate.
CC -1- PATHWAY: Second phase of glycolysis; second step.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- SIMILARITY: Belongs to the phosphoglycerate kinase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; AE013129; AAM24955.1; -
CC DR HAMAP; MF_00145; - 1.
CC DR InterPro; IPR001576; PGK.
CC DR Pfam; PF00162; PGK; 1.
CC DR ProSITE; PS00111; PGK; 1.
CC KW Transferrase; Kinase; Glycolysis; Complete proteome.
CC SQ SEQUENCE 394 AA; 42816 MW; 983EA959E240A32 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 394;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 593 GDSAAI 599
DB 351 GDSAAI 357

RESULT 61
PGK_CHLTE STANDARD; PRT; 397 AA.
AC Q8KAE1;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Phosphoglycerate kinase (EC 2.7.2.3).
GN PGK OR CT2222.
OS Chlorobium tepidum.
OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
OC Chlorobium.
OX NCBI_TaxId=1097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TUS / ATCC 49652 / DSM 12025;
RX MEDLINE=22103685; PubMed=12093901;
RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,
RA Dodson R.J., Deboy R., Gwin M.L., Nelson W.C., Haft D.H.,
RA Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,
RA Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,
RA Nietman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Raddue D.,
RA Vamathevan J., Khouri H., White O., Gruber T.M., Ketchum K.A.,
RA Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;
RT "The complete genome sequence of Chlorobium tepidum TUS, a

```

```

RT  photosynthetic, anaerobic, green-sulfur bacterium";
RL  Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).
CC  -1- CATALYTIC ACTIVITY: ATP + 3-phospho-D-glycerate = ADP + 3-
CC  phospho-D-glycerol phosphate.
CC  -1- PATHWAY: Second phase of glycolysis; second step.
CC  -1- SUBUNIT: Monomer (By similarity).
CC  -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC  -1- SIMILARITY: Belongs to the phosphoglycerate kinase family.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL: AE012967; AAM73438.1; -.
CC  DR  TIGR: CT2222; -.
CC  DR  HAMAP: MF_00145; -; 1.
CC  DR  InterPro: IPR001576; PGK.
CC  DR  Pfam: PF00162; PGK, 1.
CC  DR  PRINTS: PR00477; PHGLYCKINASE.
CC  DR  PROSITE: PS00111; PGDYCERATE_KINASE; 1.
CC  KW  transferase; Kinase; Glycolysis; Complete proteome.
CC  SQ  SEQUENCE 397 AA; 42289 MW; 375A08041F0224A7 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 397;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 593 GDSAAAI 599
    |||||
Db 355 GDSAAAI 361

RESULT 62
IFS_YEAST STANDARD; PRT; 405 AA.
AC P38f31.
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Eukaryotic translation initiation factor 5 (eIF-5).
GN TIF5 OR YPR041W OR YP1085.05.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OC NCBI_TaxID=4932;
RP SEQUENCE FROM N.A. AND SEQUENCE OF 19-30; 326-340 AND 363-377.
RC STRAIN=X2180 / ATCC 26109;
MEDLINE=93252941; PubMed=8486705;
RA Chakravarti D., Maitra U.;
RT "Eukaryotic translation initiation factor 5 from Saccharomyces
RT cerevisiae. Cloning, characterization, and expression of the gene
RT encoding the 45346-Da protein.";
RL J. Biol. Chem. 268:10524-10533(1993).
RN [1]
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=IL125-2B;
RX MEDLINE=98398050; PubMed=9730282;
RA Maciewicz-Staniorowska B., Skala J., Jasinski M., Grenson M.,
RA Goffeau A., Ulaszewska S.;
RT "Functional analysis of three adjacent open reading frames from the
RT right arm of yeast chromosome XVI.";
RL Yeast 14:1027-1039(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=9731371; PubMed=9169875;
RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Anseorge W.,
RA Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,

```

```

RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA Dietrich F.S., Dilius H., Dipolo T., Dubois E., Duesterhoeft A.,
RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.,
RA Huntje-Smith S., Hyman R., Johnston M., Kaiman S., Kleine K.,
RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
RA Marache R., Messing O., Mewes H.-W., Miripat S., Moestl D.,
RA Mueller-Auer S., Namath A., Neutwich U., Oefner P., Pearson D.,
RA Petel F.X., Pohl T.M., Purnelle D., Schaefer M., Scharfe M.,
RA Scherens B., Schramm S., Schroeder M., Scluz A.M., Tettelin H.,
RA Uristarazu L.A., Ushinsky S., Viereckels F., Vissers S., Voss H.,
RA Walsh S.V., Wambolt R., Wang Y., Wedler E., Wedler H., Winnett E.,
RA Zhong W.W., Zollner A., Vo D.H., Hani J.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";
RL Nature 387:103-105(1997).
CC -1- FUNCTION: CATALYZES THE HYDROLYSIS OF GTP BOUND TO THE 40S
CC RIBOSOMAL INITIATION COMPLEX (40S.RNA.MET-TRNA[F].EIF-2.GTP) WITH
CC THE SUBSEQUENT JOINING OF A 60S RIBOSOMAL SUBUNIT RESULTING IN THE
CC RELEASE OF EIF-2 AND THE GUANINE NUCLEOTIDE. THE SUBSEQUENT
CC JOINING OF A 60S RIBOSOMAL SUBUNIT RESULTS IN THE FORMATION OF A
CC FUNCTIONAL 80S INITIATION COMPLEX (80S.RNA.MET-TRNA[F]). EIF-5 IS
CC ESSENTIAL FOR CELL VIABILITY.
CC -1- SUBUNIT: Monomer.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative initiation;
CC Comment=2 isoforms, long (shown here) and short, are produced by
CC alternative initiation;
CC -1- SIMILARITY: BELONGS TO THE EIF-2-BETA / EIF-5 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L10840; -; NOT ANNOTATED_CDS.
CC EMBL: Z71255; CA94989.1; -.
CC DR  EMBL: Z68111; CA92145.1; -.
CC DR  EMBL: Z73616; CA97991.1; -.
CC DR  PIR: A46699; A46699.
CC DR  SCD: S006245; TIF5.
CC DR  GO: GO:0005843; C:cytosolic small ribosomal subunit (sensu Eu. . .; IPI.
CC DR  GO: GO:0005936; F:GTPase activator activity; IMP.
CC DR  GO: GO:0042256; P:mature ribosome assembly; IMP.
CC DR  GO: GO:006446; P:regulation of translational initiation; IMP.
CC DR  InterPro: IPR002735; eIF5_eIF2B.
CC DR  InterPro: IPR003307; eIF5C.
CC DR  Pfam: PF01873; eIF5_eIF2B; 1.
CC DR  Pfam: PF02020; W2; 1.
CC DR  ProDom: PD004078; eIF5_eIF2B; 1.
CC DR  SMART: SM00653; eIF2B_5; 1.
CC DR  SMART: SM00515; eIF5C; 1.
CC KW  Initiation factor; Protein biosynthesis; Alternative initiation;
CC GTP-binding.
FT CHAIN 1 405 EUKARYOTIC TRANSLATION INITIATION FACTOR
FT FT 5, ISOFORM LONG.
FT FT EUKARYOTIC TRANSLATION INITIATION FACTOR
FT FT 5, ISOFORM SHORT.
FT FT FOR ISOFORM SHORT.
FT NP_BIND 152 156 POLY-LYS.
FT FT 27 34 GTP (POTENTIAL).
FT FT 396 405 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 405 AA; 45261 MW; 1A1DA5634ED1B1F CRC64;

Query Match 0.9%; Score 7; DB 1; Length 405;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 94 DEDIVNE 100
    |||||

```

Db 297 DEDIVNE 303

```

RESULT 63
RPA3_YEAST STANDARD; PRT; 415 AA.
ID_RPA3_YEAST
AC 001080;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE DNA-directed RNA polymerase I 49 kDa polypeptide (EC 2.7.7.6) (A49).
GN RPA49 OR RNRI3 OR YNU248C OR N0880.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Saccharomycetes; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CMY214;
RC MEDLINE=93028452; PubMed=1409638;
RA Liljelund P., Mariotte S., Buhler J.-M., Sentenac A.;
RT "Characterization and mutagenesis of the gene encoding the A49
RT subunit of RNA polymerase I in Saccharomyces cerevisiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:9302-9305(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=5288c / FY1679;
RC MEDLINE=97377992; PubMed=9234673;
RA Sen-Gupta M., Gueldeuer U., Beinhauer J.D., Fiedler T.A.,
RA Hegemann J.H.;
RT "Sequence analysis of the 33 kb long region between ORC5 and SU11
RT from the left arm of chromosome XIV from Saccharomyces cerevisiae.";
RL Yeast 13:849-860(1997).
CC -I- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES. RNA POLYMERASE A IS ESSENTIALLY USED TO TRANSCRIBE
CC RIBOSOMAL DNA UNITS.
CC -I- FUNCTION: A49 IS EASILY DISSOCIATED FROM THE REST OF POL A,
CC PRODUCING THE FORM A', WHICH SHOWS IMPAIRED TRANSCRIPTIONAL
CC ACTIVITY AND INCREASED SENSITIVITY TO ALPHA-AMANTIN. THE FUNCTION
CC OF A49 MIGHT BE LINKED TO THE RNASE H ACTIVITY THAT WAS FOUND
CC ASSOCIATED WITH THIS SUBUNIT.
CC -I- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC -I- SUBUNIT: RNA POLYMERASE I CONSISTS OF 14 DIFFERENT SUBUNITS.
CC -I- SUBCELLULAR LOCATION: Nuclear.
CC -I- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
CC III FOR 5S AND TRNA GENES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M96600; AAA34380.1; -
CC EMBL; X96722; CAA65496.1; -
CC EMBL; Z71524; CAA96155.1; -
CC PIR; S63221; S63221
CC SGD; S0005192; RPA49.
KW Transferase; DNA-directed RNA polymerase; Transcription;
KW Nuclear protein.
FT CONFLICT 66 T -> P (IN REF. 1).
FT CONFLICT 157 S -> C (IN REF. 1).
FT SEQUENCE 415 AA; 46650 MW; 3DBF05440D26021 CRC64;
SQ

```

QY 469 KLLERLN 475
 Db 282 KLLERLN 288

```

RESULT 64
HEM1_AERPE STANDARD; PRT; 416 AA.
ID_HEM1_AERPE
AC 093372;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glutamyl-tRNA reductase (EC 1.2.1.-) (GLUTR).
GN HEMA OR APE2296.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcaceae;
OC Desulfurococcaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KI;
RC MEDLINE=9310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix KI.";
RL DNA Res. 6:83-101(1999).
CC -I- CATALYTIC ACTIVITY: Glutamyl-tRNA(Glu) + NADPH = glutamate-1-
CC semialdehyde + NADP(4+) + tRNA(Glu).
CC -I- PATHWAY: Porpyrin biosynthesis by the C5 pathway; first step.
CC -I- SIMILARITY: Belongs to the glutamyl-tRNA reductase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration -
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP000064; BAA81308.1; -
CC PIR; D72456; D72456.
CC HAMAP; MF_00087; -; 1.
CC InterPro; IPR000343; GLUTR.
CC Pfam; PF00745; GLUTR dimer; 1.
CC Pfam; PF05201; GLUTR_N; 1.
CC Pfam; PF05200; GLUTR_NAD_bind; 1.
CC TIGRPFAM; TIGR01035; hema; 1.
CC PROSITE; PS00747; GLUTR; 1.
KW Porpyrin biosynthesis; Oxidoreductase; NADP. Complete proteome.
FT ACT_SITE 47 47 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 87 87 BASE (BY SIMILARITY).
FT SEQUENCE 416 AA; 45244 MW; 6E087B36FC6853D CRC64;
SQ

```

QY 753 TLAGLN 759
 Db 346 TLAGLN 352

RESULT 65
 Y12A_ECOLI STANDARD; PRT; 424 AA.
 ID_Y12A_ECOLI
 AC P51026; Q47535;
 DT 01-OCT-1996 (Rel. 34, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Insertion element IS2A hypothetical 48.2 kda protein.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RA Nishimoto H., Saito N.;
 RL Submitted (JUN-1996) to the EMBL/Genbank/DBJ databases.
 CC -! SIMILARITY: BELONGS TO THE TRANSPOSASE FAMILY 8.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; D85613; BAA12836.1; -
 CC DR PIR; S78610; S78610.
 CC DR InterPro; IPR001584; Rve.
 CC DR InterPro; IPR002514; Transposase_8.
 CC DR Pfam; PF01527; Transposase_8; 1.
 CC DR Pfam; PF00665; rve; 1.
 CC DR Transposable element; Hypothetical protein.
 CC KW SEQUENCE 424 AA; 48153 MW; A791CE858D05A329 CRC64;
 SQ

Query Match 0.9%; Score 7; DB 1; Length 424;
 Best Local Similarity 100.0%; Pred. No. 94;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 202 ASELAAA 208
 DB 85 ASELAAA 91

RESULT 66
 TRPB_HALN1 STANDARD; PRT; 430 AA.
 AC Q9HSC0;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Tryptophan synthase beta chain (EC 4.2.1.20).
 GN TRPB OR VNG0307G.
 OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
 OC Archaea; Euryarchaeota; Halobacteriia; Halobacteriales;
 OC Halobacteriaceae; Halobacterium.
 OX NCBI_TaxID=64091;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20504483; PubMed=11016950;
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M.,
 RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Geo Y.A.,
 RA Leitbauer B., Keiler K., Cruz R., Danson M.J., Hough D.W.,
 RA Maddock D.G., Jablonski P.E., Krebs M.P., Angewill C.M., Dale H.,
 RA Isenbarger T.A., Peck R.F., Pohlischer M., Spidichler J.L., Jung K.-H.,
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
 RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
 RL "Genome sequence of Halobacterium species NRC-1."
 CC Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
 CC -! FUNCTION: The beta subunit is responsible for the synthesis of L-
 CC tryptophan from indole and L-serine.
 CC -! CATALYTIC ACTIVITY: L-serine + 1-(indol-3-yl)glycerol 3-phosphate
 CC = L-tryptophan + glyceraldehyde 3-phosphate.
 CC -! COFACTOR: Pyridoxal phosphate (By similarity).
 CC -! PATHWAY: Tryptophan biosynthesis; fifth (last) step.
 CC -! SUBUNIT: Tetramer of two alpha and two beta chains (By

CC similarity).
 CC -! SIMILARITY: BELONGS TO THE TRPB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; AE004991; AAC8887.1; -
 CC DR PIR; C84190; C84190.
 CC DR HSSP; P00933; 2MSY.
 CC DR HAMAP; MF_00133; -; 1.
 CC DR InterPro; IPR001926; B6_enzyme_beta.
 CC DR InterPro; IPR006653; TYP_synth_beta.
 CC DR InterPro; IPR006654; TYP_synth_beta.
 CC DR Pfam; PF00291; PALP; 1.
 CC DR TIGRFAMs; TIGR00263; trpb; 1.
 CC DR PROSITE; PS00168; TRP_SYNTHASE_BETA; 1.
 CC DR Tryptophan biosynthesis; Pyridoxal phosphate; lyase;
 CC KW Complete proteome.
 CC KM BINDING 95
 CC SQ SEQUENCE 430 AA; 45451 MW; B1B56DAC9AD0BCDC CRC64;
 SQ

Query Match 0.9%; Score 7; DB 1; Length 430;
 Best Local Similarity 100.0%; Pred. No. 95;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 AIISEEL 83
 DB 214 AIISEEL 220

RESULT 67
 GSA_AERPE STANDARD; PRT; 432 AA.
 ID GSA_AERPE
 AC Q9Y3I9;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable glutamate-1-semialdehyde 2,1-aminomutase (EC 5.4.3.8) (GSA)
 DE (Glutamate-1-semialdehyde aminotransferase) (GSA-AT).
 GN HEMT OR APE2299.
 OS Aeropyrum pernix.
 OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
 OC Desulfurococcaceae; Aeropyrum.
 OX NCBI_TaxID=56636;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K1;
 RX MEDLINE=99310339; PubMed=10382966;
 RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
 RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Akai A., Kosugi H.,
 RA Hoshino A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
 RA Yamazaki Y., Kusuda N., Oguchi A., Aoki K.-I., Kubota K.,
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
 RT "Complete genome sequence of an aerobic hyper-thermophilic
 RT crenarchaeon, Aeropyrum pernix K1."
 RL DNA Res. 6:83-101(1999).
 CC -! CATALYTIC ACTIVITY: (S)-4-amino-5-oxopentanoate = 5-
 CC aminovalinate.
 CC -! COFACTOR: Pyridoxal phosphate.
 CC -! PATHWAY: Porphyrin biosynthesis by the C5 pathway; second step.
 CC -! SUBCELLULAR LOCATION: Cytoplasmic (potential).
 CC -! SIMILARITY: Belongs to class-III of pyridoxal-phosphate-dependent
 CC aminotransferases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: A000064; BAA81311.1; -
DR PIR: G72456; G72456.
DR HSSP: P24630; 2GSA.
DR HAMAP: MF_00375; -; 1.
DR InterPro: IPR005814; AminoTrans_3.
DR InterPro: IPR004639; HemL.
DR Pfam: PF0202; aminoTran_3; 1.
DR TIGRPFAM: TIGR00713; hemL; 1.
DR PROSITE: PS00600; AA_TRANSFERRIN_CLASS_3; 1.
DR Porphyryn biosynthesis; Isomerase; Pyridoxal phosphate;
KW Complete proteome.
FT BINDING 269 269 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 432 AA; 46281 MW; A8FC01657A163BD2 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 432;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 733 EAEVLLA 739
DB 94 EAEVLLA 100

RESULT 68
TBG_SCHUP STANDARD; PRT; 446 AA.

AC 09Y882;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Tubulin gamma chain (Gamma tubulin).
GN TUG1 OR GTB1.
OS Schizosaccharomyces japonicus.
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4897;
RP SEQUENCE FROM N.A.
RT Horio T., Shimizu M., Akashi T., Tanaka K.;
RT "Isolation and characterization of gamma-tubulin gene from
RT Schizosaccharomyces japonicus."
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: TUBULIN IS THE MAJOR CONSTITUENT OF MICROTUBULES.
CC GAMMA TUBULIN IS FOUND AT MICROTUBULE ORGANIZING CENTERS (MTOC)
CC SUCH AS THE SPINDLE POLES OR THE CENTROSOME, SUGGESTING THAT IT
CC IS INVOLVED IN THE MINUS-END NUCLEATION OF MICROTUBULE ASSEMBLY.
CC -!- SIMILARITY: BELONGS TO THE TUBULIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AF159163; AAD41900.1; -
DR InterPro: IPR000217; Tubulin.
DR InterPro: IPR003008; Tubulin_FtsZ.
DR Pfam: PF00091; tubulin; 1.
DR Pfam: PF03953; tubulin_C; 1.
DR PRINTS: PRO1161; TUBULIN.
DR PROSITE: PS00227; TUBULIN; 1.
KW Microtubules; GTP-binding.
FT NP_BIND 142 148 GTP (POTENTIAL).
SQ SEQUENCE 446 AA; 50158 MW; B86BBE95A94A2C CRC64;

Query Match 0.9%; Score 7; DB 1; Length 446;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 470 LIERLND 476
DB 153 LIERLND 159

RESULT 69
TBG_SCHPO STANDARD; PRT; 446 AA.

AC P25295;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Tubulin gamma chain (Gamma tubulin).
GN TUG1 OR GTB1 OR SPBC32F12.04.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RP SEQUENCE FROM N.A.
RT Stearns T., Evans L., Kirschner M.;
RT "Gamma-tubulin is a highly conserved component of the centrosome."
RL Cell 65:825-836(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92121241; PubMed=1770000;
RX Horio T., Uzawa S., Jung M.K., Oakley B.R., Tanaka K., Yanagida M.;
RT "The fission yeast gamma-tubulin is essential for mitosis and is
RT localized at microtubule organizing centers."
RL J. Cell Sci. 99:693-700(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RX Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
RA Rutherford K., Rutter S., Saunders J., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grynpres B.,
RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Meller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Meestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Medler H., Wandut R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Useery D., Barrett B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).

CC -!- FUNCTION: TUBULIN IS THE MAJOR CONSTITUENT OF MICROTUBULES.
CC GAMMA TUBULIN IS FOUND AT MICROTUBULE ORGANIZING CENTERS (MTOC)
CC SUCH AS THE SPINDLE POLES OR THE CENTROSOME, SUGGESTING THAT IT
CC IS INVOLVED IN THE MINUS-END NUCLEATION OF MICROTUBULE ASSEMBLY.
CC -!- SIMILARITY: BELONGS TO THE TUBULIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M63447; AAA35305.1; -
 DR EMBL: X62031; CAA3976.1; -
 DR EMBL: AL023796; CAA19365.1; -
 DR PIR: B39528; UBZPG.
 DR GeneDB: SPOMB; SPBC32F12.04; -
 DR InterPro: IPR000217; TUBULIN.
 DR InterPro: IPR003008; TUBULIN_FtsZ.
 DR Pfam: PF00091; tubulin_1.
 DR Pfam: PF03953; TUBULIN_C_1.
 DR PRINTS: PR01161; TUBULIN.
 DR PROSITE: PS00227; TUBULIN; 1.
 KW Microtubules; GTP-binding.
 FT NP BIND 142 RA 148 GTP (POTENTIAL).
 SO SEQUENCE 446 AA; 49965 MW; 8A48833EAB3BD469 CRC64;
 Query Match 0.9%; Score 7; DB 1; Length 446;
 Best Local Similarity 100.0%; Pred. No. 99;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 470 LIERAND 476
 Db 153 LIERAND 159
 RESULT 70
 V50K_BYDVP
 ID V50K_BYDVP STANDARD; PRT; 450 AA.
 AC P09516;
 DT 01-MAR-1989 (Rel. 10; Created)
 DT 28-FEB-2003 (Rel. 41; Last sequence update)
 DT 28-FEB-2003 (Rel. 41; Last annotation update)
 DE 50 kDa protein (ORF 4).
 OS Barley yellow dwarf virus (isolate PAV) (BYDV).
 OS Virusae; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;
 OC Luteovirus.
 OC NCB1_TaxID=12040;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88289355; PubMed=3399386;
 RA Miller W.A., Waterhouse P.M., Gerlach W.L.;
 RT "Sequence and organization of barley yellow dwarf virus genomic RNA";
 RL Nucleic Acids Res. 16:6097-6111 (1988).
 CC -1- SIMILARITY: TO ORF 6 OF BEET WESTERN YELLOWS VIRUS. ORF6 OF
 CC POTATO LEAFROLL VIRUS AND OF 54K PROTEIN OF BEET NECROTIC
 CC YELLOW VEIN VIRUS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X07653; CAA30495.2; -
 DR InterPro: IPR002929; PLRV_ORFS.
 DR Pfam: PF01690; PLRV_ORFS; 1.
 SO SEQUENCE 450 AA; 49765 MW; 3960B263D45331D CRC64;
 Query Match 0.9%; Score 7; DB 1; Length 450;
 Best Local Similarity 100.0%; Pred. No. 99;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 351 POUTPEP 357
 Db 351 POUTPEP 357

Db 24 POUTPEP 30
 RESULT 71
 TBGI_HUMAN
 ID TBGI_HUMAN STANDARD; PRT; 451 AA.
 AC P23258; Q9BM59;
 DT 01-NOV-1991 (Rel. 20; Created)
 DT 28-FEB-2003 (Rel. 41; Last sequence update)
 DT 15-SEP-2003 (Rel. 42; Last annotation update)
 DE Tubulin gamma-1 chain (Gamma-1 tubulin) (Gamma-tubulin complex
 DE component 1) (GCP-1).
 GN TBGI OR TUBG.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OK NCB1_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91249388; PubMed=1904010;
 RA Zheng Y., Jung M.K., Oakley B.R.;
 RT "Gamma-tubulin is present in Drosophila melanogaster and Homo sapiens
 RT and is associated with the centrosome";
 RL Cell 65:817-823(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Skin;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stappleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
 RA Brownstein M.J., Usdin J.B., Tomihyori S., Carninci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultk S.W.,
 RA Villalón D.K., Muzny D.C., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Boultard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
 RA Schercher A., Schein J.E., Jones S.J.M., Matra M.A.;
 RA "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RP MASS SPECTROMETRY.
 RC TISSUE=Breast cancer;
 RX MEDLINE=21829512; PubMed=11840567;
 RA Harris R.A., Yang A., Stein R.C., Lucy K., Brusten L., Herath A.,
 RA Parekh R., Waterfield M.D., O'Hare M.J., Neville M.A., Page M.J.,
 RA Zvelebil M.J.;
 RT "Cluster analysis of an extensive human breast cancer cell line
 RT protein expression map database";
 RL Proteomics 2:212-223(2002).
 RN [4]
 RP SUBCELLULAR LOCATION.
 RX MEDLINE=98234404; PubMed=9566969;
 RA Tassin A.-M., Celati C., Moudjou M., Bornens M.;
 RT "Characterization of the human homologue of the yeast spc98p and its
 RT association with gamma-tubulin";
 RL J. Cell Biol. 141:689-701(1998).
 CC -1- FUNCTION: TUBULIN IS THE MAJOR CONSTITUENT OF MICROTUBULES.
 CC SUCH AS THE SPINDLE POLES OR THE CENTROSOME, SUGGESTING THAT IT
 CC IS INVOLVED IN THE MINUS-END NUCLEATION OF MICROTUBULE ASSEMBLY.
 CC -1- SUBUNIT: Interacts with GCP2 and GCP3.
 CC -1- SUBCELLULAR LOCATION: Centrosome.
 CC -1- MASS SPECTROMETRY: MW=51197.98; METHOD=MALDI.
 CC -1- SIMILARITY: BELONGS TO THE TUBULIN FAMILY.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M61764; AAA52620.1; -
DR EMBL; BC000619; AAH00619.1; -
DR PIR; A39527; UBHUG.
DR Genew; HGNC:12417; TUBG1.
DR MIM; 191135; -
DR GO; GO:0005813; C:centrosome; TAS.
DR GO; GO:0005874; C:microtubule; TAS.
DR GO; GO:0005816; C:spindle pole body; TAS.
DR GO; GO:0005200; F:structural constituent of cytoskeleton; TAS.
DR GO; GO:0000226; P:microtubule cytoskeleton organization and b. .; TAS.
DR InterPro; IPR000217; Tubulin.
DR InterPro; IPR003008; Tubulin_Fcsz.
DR Pfam; PF00091; tubulin_1.
DR Pfam; PF03953; tubulin_C; 1.
DR PRINTS; PR01161; TUBULIN.
DR PROSITE; PS00227; TUBULIN; 1.
KW Microtubules; GTP-binding.
KW NP BIND 142 148 GTP (POTENTIAL).
FT CONFLICT 202 202 V -> A (IN REF. 1).
FT CONFLICT 202 202 V -> L (IN REF. 1).
SQ SEQUENCE 451 AA; 51170 MW; E2A4C0179ED0CFE8 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 1e+02; Length 451;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 470 LIERLND 476
DB 153 LIERLND 159

RESULT 72
TBG1 MOUSE STANDARD; PRT; 451 AA.
ID TBG1 MOUSE
AC 092310;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Tubulin gamma-1 chain (Gamma-1 tubulin) (Gamma-tubulin complex
component 1) (GCP-1).
GN TUBG1 OR TUBG.
OS Mus musculus (Mouse). and
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCB1_TaxID=10090, 10116;
OX [1]
RP SEQUENCE FROM N.A.
RC SPECTES=Mouse;
RX MEDLINE=22388257; PubMed=12477932;
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schaller G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.J., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Locquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodegriem E.J., Lu X., Gibbs S.A.,
RA Fahy J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

```

```

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Maira M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECTES=Rat;
RX MEDLINE=99397453; PubMed=10470852;
RA Nakada T., Okada N., Makino Y., Tamura T.;
RT "Structure of rat gamma-tubulin and its binding to Hp33.";
RL DNA Res. 6:207-209(1999).
CC -1- FUNCTION: Tubulin is the major constituent of microtubules. Gamma
CC tubulin is found at microtubule organizing centers (MTOC) such as
CC the spindle poles or the centrosome, suggesting that it is
CC involved in the minus-end nucleation of microtubule assembly (By
CC similarity).
CC -1- SUBUNIT: Interacts with GCP2 and GCP3 (By similarity).
CC -1- SUBCELLULAR LOCATION: Centrosome (By similarity).
CC -1- SIMILARITY: BELONGS TO THE TUBULIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; BC006581; AAH06581.1; -
DR EMBL; AB015946; BAA36504.1; -
DR MGD; MGI:101834; Tubb.
DR InterPro; IPR000217; Tubulin.
DR InterPro; IPR003008; Tubulin_Fcsz.
DR Pfam; PF00091; tubulin_1.
DR Pfam; PF03953; tubulin_C; 1.
DR PRINTS; PR01161; TUBULIN.
DR PROSITE; PS00227; TUBULIN; 1.
KW Microtubules; GTP-binding.
KW NP BIND 142 148 GTP (POTENTIAL).
FT CONFLICT 451 AA; 51100 MW; A8F1068D12D0C88A CRC64;
SQ SEQUENCE 451 AA; 51100 MW; A8F1068D12D0C88A CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 1e+02; Length 451;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 470 LIERLND 476
DB 153 LIERLND 159

RESULT 73
TBG2 HUMAN STANDARD; PRT; 451 AA.
ID TBG2 HUMAN
AC 09NR33;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Tubulin gamma-2 chain (Gamma-2 tubulin).
GN TUBG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCB1_TaxID=9606;
OX [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20366133; PubMed=10903841;
RA Wise D.O., Krahe R., Oakley B.R.;
RT "The gamma-tubulin gene family in humans.";
RL Genomics 67:164-170(2000).
RN [2]
RP SEQUENCE FROM N.A.

```

RC TISSUE=Mammary gland;
 RA Iisoeai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shitatori A., Sudo H., Sugawara M.,
 RA Watanabe M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Watanabe S., Kimura K., Murakami K., Ishi S., Kawai Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y.,
 RA Niimiya K., Iwayanagi T.;
 RT "NEO human cDNA sequencing project";
 RL Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.

RN [3]
 RP SEQUENCE FROM N.A.

RC TISSUE=Lung;
 RX MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.A., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

CC -I- FUNCTION: Tubulin is the major constituent of microtubules. Gamma
 CC tubulin is found at microtubule organizing centers (MTOC) such as
 CC the spindle poles or the centrosome, suggesting that it is
 CC involved in the minus-end nucleation of microtubule assembly.
 CC -I- SIMILARITY: BELONGS TO THE TUBULIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; AF225971; AAF34188.1; -
 CC EMBL; AK022324; BAB14012.1; -
 CC EMBL; BC009670; AAH09670.1; -
 CC Genbank; HGNC:12419; TUBG2.
 DR MIM; 605785; -
 DR GO; GO:0005198; F:structural molecule activity; TMS.
 DR InterPro; IPR000217; Tubulin.
 DR InterPro; IPR003008; Tubulin_FtsZ.
 DR Pfam; PF00091; tubulin_1.
 DR Pfam; PF03953; tubulin_C_1.
 DR PRINTS; PR01161; TUBULIN.
 DR PROSITE; PS00227; TUBULIN; 1.
 KW Microtubules; GTP-binding.
 FT NP BIND 142 148
 FT SEQUENCE 451 AA; 51091 MW; FDE2CB5A33D92691 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 451;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 470 LLERLND 476
 Db 153 LLERLND 159

RESULT 74
 TBG2 MOUSE STANDARD; PRT: 451 AA.
 AC O8VCK3;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Tubulin gamma-2 chain (Gamma-2 tubulin).
 GN TUBG2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RP [1]
 RP SEQUENCE FROM N.A.

RC TISSUE=Retina;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.A., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

CC -I- FUNCTION: TUBULIN IS THE MAJOR CONSTITUENT OF MICROTUBULES.
 CC GAMMA TUBULIN IS FOUND AT MICROTUBULE ORGANIZING CENTERS (MTOC)
 CC SUCH AS THE SPINDLE POLES OR THE CENTROSOME, SUGGESTING THAT IT
 CC IS INVOLVED IN THE MINUS-END NUCLEATION OF MICROTUBULE ASSEMBLY.
 CC -I- SIMILARITY: BELONGS TO THE TUBULIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; BC019652; AAH19652.1; -
 CC InterPro; IPR000217; Tubulin.
 CC InterPro; IPR003008; Tubulin_FtsZ.
 DR Pfam; PF00091; tubulin_1.
 DR Pfam; PF03953; tubulin_C_1.
 DR PRINTS; PR01161; TUBULIN.
 DR PROSITE; PS00227; TUBULIN; 1.
 KW Microtubules; GTP-binding.
 FT NP BIND 142 148
 FT SEQUENCE 451 AA; 51121 MW; JPC0CF1E04839A CRC64;

Query Match 0.9%; Score 7; DB 1; Length 451;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 470 LLERLND 476
 Db 153 LLERLND 159

RESULT 75
 TBG_XENLA


```

ID TBG_XENLA STANDARD; PRT; 451 AA.
AC P23330;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
GN XGAM.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodidae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91249389; PubMed=1840506;
RA Stearns T., Evans L., Kirschner M.;
RT "Gamma-tubulin is a highly conserved component of the centrosome";
RL Cell 65:825-836(1991).
CC -1- FUNCTION: TUBULIN IS THE MAJOR CONSTITUENT OF MICROTUBULES.
CC GAMMA TUBULIN IS FOUND AT MICROTUBULE ORGANIZING CENTERS (MTOC)
CC SUCH AS THE SPINDLE POLES OR THE CENTROSOME, SUGGESTING THAT IT
CC IS INVOLVED IN THE MINUS-END NUCLEATION OF MICROTUBULE ASSEMBLY.
CC -1- SIMILARITY: BELONGS TO THE TUBULIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M63446; AAA49720.1; -
DR PIR; A39528; UBXLG.
DR InterPro; IPR000217; Tubulin.
DR InterPro; IPR003008; Tubulin_fsz.
DR Pfam; PF00091; tubulin_1.
DR Pfam; PF03953; tubulin_C_1.
DR PRINTS; PR01161; TUBULIN.
DR PROSITE; PS00227; TUBULIN; 1.
DR Microtubules; GTP-binding.
KW NP_BIND 142 148 GTP (POTENTIAL).
FT SEQUENCE 451 AA; 51167 MW; 52195DA5921EF246 CRC64;
SQ

```

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 451;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 470 LTERLND 476
Db 153 LTERLND 159

```

RESULT 76
ID TBG_XENLA STANDARD; PRT; 455 AA.
AC P23330;
DT 01-NOV-1991 (Rel. 35, Created)
DT 01-NOV-1991 (Rel. 35, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
GN XGAM.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodidae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91249389; PubMed=1840506;
RA Stearns T., Evans L., Kirschner M.;
RT "Gamma-tubulin is a highly conserved component of the centrosome";
RL Cell 65:825-836(1991).
CC -1- FUNCTION: TUBULIN IS THE MAJOR CONSTITUENT OF MICROTUBULES.
CC GAMMA TUBULIN IS FOUND AT MICROTUBULE ORGANIZING CENTERS (MTOC)
CC SUCH AS THE SPINDLE POLES OR THE CENTROSOME, SUGGESTING THAT IT
CC IS INVOLVED IN THE MINUS-END NUCLEATION OF MICROTUBULE ASSEMBLY.
CC -1- SIMILARITY: BELONGS TO THE TUBULIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M63446; AAA49720.1; -
DR PIR; A39528; UBXLG.
DR InterPro; IPR000217; Tubulin.
DR InterPro; IPR003008; Tubulin_fsz.
DR Pfam; PF00091; tubulin_1.
DR Pfam; PF03953; tubulin_C_1.
DR PRINTS; PR01161; TUBULIN.
DR PROSITE; PS00227; TUBULIN; 1.
DR Microtubules; GTP-binding.
KW NP_BIND 142 148 GTP (POTENTIAL).
FT SEQUENCE 451 AA; 51167 MW; 52195DA5921EF246 CRC64;
SQ

```

CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U41028; AAA82357.1; -
DR PIR; T15622; T15622.
DR Wormpep; C25G6.5; CE04086.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCR_RHODOPSIN.
DR PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR PROSITE; PS00262; G_PROTEIN_RECEP_F1_2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein.
FT DOMAIN 1 45 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 46 66 1 (POTENTIAL).
FT DOMAIN 67 86 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 87 107 2 (POTENTIAL).
FT DOMAIN 108 120 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 121 135 3 (POTENTIAL).
FT DOMAIN 136 152 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 153 173 4 (POTENTIAL).
FT DOMAIN 174 210 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 211 231 5 (POTENTIAL).
FT DOMAIN 232 272 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 273 293 6 (POTENTIAL).
FT DOMAIN 294 307 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 308 328 7 (POTENTIAL).
FT DOMAIN 329 455 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 5 5 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 29 29 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 455 AA; 52187 MW; CIDP3C3C2E2EC65F5 CRC64;

```

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 455;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 246 TTNNTS 252
Db 250 TTNNTS 256

```

RESULT 77
ID SELA_PASMU STANDARD; PRT; 460 AA.
AC Q9CK66;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE L-seryl-tRNA(Sec) selenium transferase (EC 2.9.1.1) (Cysteiny]-
DE tRNA(Sec) selenium transferase (Selenocysteine synthase)
DE (Selenocysteinyl-CRNA (Sec) synthase).
GN SELA OR PM1768.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21145666; PubMed=11248100;
RA May B.J., Zhang O., Li L.L., Paustian M.L., Whitlam T.S., Kapur V.;
RT Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -1- FUNCTION: Converts seryl-tRNA(Sec UCA) to selenocysteinyl-tRNA(Sec
CC UCA) during selenoprotein biosynthesis (By similarity).
CC -1- CATALYTIC ACTIVITY: L-seryl-tRNA(Sec) + selenophosphate = L-
CC selenocysteinyl-tRNA(Sec) + H(2)O + phosphate.

```

CC -1- COACTOR: Pyridoxal phosphate (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE SEDA FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its use
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR HAMAP; MF_00423; -.
DR InterPro; IPR004534; SEDA.
DR Pfam; PF03841; SEDA; 1.
DR TIGRFAMs; TIGR00474; sEDA; 1.
DR Transferrase; Pyridoxal phosphate; Selenium; Complete proteome.
FW BINDING 293 293 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 460 AA; 50667 MW; FFP6AB86AA0BCDD7 CRC64;
-----
Query Match 0.9%; Score 7; DB 1; Length 460;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 266 DIDSLLK 272
Db 449 DIDSLLK 455
-----
RESULT 78
TBG1_EUPOC STANDARD; PRT; 461 AA.
ID TBG1_EUPOC
AC P34786;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Tubulin gamma-1 chain (Gamma-1 tubulin).
OS Euplotes octocarinatus.
OC Eukaryota; Alveolata; ciliophora; Spirotrichea; Hypotrichia;
OC Euplotida; Euplotidae; Euplotes.
OX NCBI_TaxID=5937;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=3(58) -IX;
RX MEDLINE=94124022; PubMed=8294024;
RA Liang A., Heckmann K.;
RT "The macronuclear gamma-tubulin-encoding gene of Euplotes
RT octocarinatus contains two introns and an in-frame TGA.";
RN Gene 136:319-322(1993).
[2]
RN REVISIONS.
RA Liang A.;
CC -1- FUNCTION: (OCT-1995) to the EMBL/GenBank/DBJ databases.
CC GAMMA TUBULIN IS FOUND AT MICROTUBULE ORGANIZING CENTERS (MTOC)
CC SUCH AS THE SPINDLE POLES OR THE CENTROSOME, SUGGESTING THAT IT
CC IS INVOLVED IN THE MINDS-END NUCLEATION OF MICROTUBULE ASSEMBLY.
CC -1- SIMILARITY: BELONGS TO THE TUBULIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its use
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X71353; CAA50488.1; -.
DR EMBL; Y09552; CAA70743.1; -.
DR EMBL; Y09554; CAA70745.1; -.
DR InterPro; IPR000217; Tubulin.
DR InterPro; IPR003008; Tubulin_Psz.

```

```

DR Pfam: PF00091; tubulin; 1.
DR Pfam: PF03953; tubulin; C; 1.
DR PRINTS: PR01161; TUBULIN.
DR PROSITE: PS00227; TUBULIN; 1.
DR Microcubules; GTP-binding.
FT NP BIND 142 148 GTP (POTENTIAL).
SQ SEQUENCE 461 AA; 51992 MW; 28029859BB939C36 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 461;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

OY 470 LLERLND 476
|||
Db 153 LLERLND 159

RESULT 79
TBG2_EUPOC STANDARD; PRT; 461 AA.
ID TBG2_EUPOC
AC P90548;
DC 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Tubulin gamma-2 chain (Gamma-2 tubulin).
OS Euplotes octocarinatus.
OC Eukaryota; Alveolata; Ciliophora; Spirotrichea; Hypotrichia;
OC Euplotida; Euplotidae; Euplotes.
OX NCBI_TaxID=5937;
RN [1]
RP SEQUENCE FROM N.A.
RA Tan M., Liang A., Heckmann K.;
RT "The two gamma tubulin genes of Euplotes octocarinatus code for a
RT slightly different protein.";
RL Submitted (May-1998) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: TUBULIN IS THE MAJOR CONSTITUENT OF MICROTUBULES.
CC SUCH AS THE SPINDLE POLES OR THE CENTROSOME, SUGGESTING THAT IT
CC IS INVOLVED IN THE MINDS-END NUCLEATION OF MICROTUBULE ASSEMBLY.
CC -I- SIMILARITY: BELONGS TO THE TUBULIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/).license@isb-sib.ch.
CC -----
DR EMBL: Y09553; CAAT0744.1; -
DR EMBL: Y17254; CAA76714.1; -
DR InterPro: IPR000217; Tubulin.
DR InterPro: IPR003008; Tubulin_FtsZ.
DR Pfam: PF00091; tubulin; 1.
DR Pfam: PF03953; tubulin; C; 1.
DR PRINTS: PR01161; TUBULIN.
DR PROSITE: PS00227; TUBULIN; 1.
DR Microcubules; GTP-binding.
FT NP BIND 142 148 GTP (POTENTIAL).
SQ SEQUENCE 461 AA; 51952 MW; DBED624C4B824A46 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 461;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

OY 470 LLERLND 476
|||
Db 153 LLERLND 159

RESULT 80
TBG2_NEUCR STANDARD; PRT; 461 AA.
ID TBG2_NEUCR

```

AC P53377;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Tubulin gamma chain (Gamma tubulin).
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
 NCBI_TaxId=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98019102; PubMed=9358070;
 RA Heckmann S., Schliwa M., Kube-Granderah E.;
 RT "Primary structure of Neurospora crassa gamma-tubulin.";
 RL Gene 199:303-309(1997).
 CC - FUNCTION: TUBULIN IS THE MAJOR CONSTITUENT OF MICROTUBULES.
 CC SUCH AS THE SPINDLE POLES OR THE CENTROSOME, SUGGESTING THAT IT
 CC IS INVOLVED IN THE MINUS-END NUCLEATION OF MICROTUBULE ASSEMBLY.
 CC - SIMILARITY: BELONGS TO THE TUBULIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X87753; CAA66348.1;
 CC InterPro; IPR000217; Tubulin.
 CC InterPro; IPR003008; Tubulin_Fte2.
 DR Pfam; PF00091; tubulin; 1.
 DR Pfam; PF03953; tubulin; C. 1.
 DR PRINTS; PR01161; TUBULIN.
 DR PROSITE; PS00227; TUBULIN; 1.
 KW Microtubules; GTP-binding.
 FT NP BIND 142 148 GTP (POTENTIAL).
 SQ SEQUENCE 461 AA; 51598 MW; 70DCCBEA224B6343 CRC64;
 Query Match 0.9%; Score 7; DB 1; Length 461;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 470 LIERLND 476
 Db 153 LIERLND 159
 RESULT 81
 TR1B_HUMAN STANDARD; PRT; 461 AA.
 AC P20333; Q16042; Q9U1H1;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Tumor necrosis factor receptor superfamily member 1B precursor (Tumor
 DE necrosis factor receptor 2) (p80) (TNF-R2) (p75) (CD120b) (Etanercept)
 DE [Contains: Tumor necrosis factor binding protein 2 (TNFR1)].
 GN TNFRSF1B OR TNFR2 OR TNFR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eumetazoa;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90260639; PubMed=2160731;
 RA Smith C.A., Davis T., Anderson D., Solam L., Beckmann M.P., Jerzy R.,
 RA Dower S.K., Cosman D., Goodwin R.G.;
 RT "A receptor for tumor necrosis factor defines an unusual family of
 RT cellular and viral proteins.";
 RL Science 248:1019-1023(1990).
 RN [2]

RP SEQUENCE FROM N.A., AND VARIANT ARG-196.
 RX MEDLINE=9104591; PubMed=2172983;
 RA Kohno T., Brewer M.T., Baker S.L., Schwartz P.E., King M.W.,
 RA Hale K.K., Squires C.H., Thompson R.C., Vannice J.L.;
 RT "A second tumor necrosis factor receptor gene product can shed a
 RT naturally occurring tumor necrosis factor inhibitor.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:8331-8335(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9629745; PubMed=8661109;
 RA Beltinger C.P., White P.S., Maris J.M., Sulman E.P., Jensen S.J.,
 RA Lepelletier D., Stallard B.J., Goeddel D.V., Desauvage F.J.;
 RA Brodeur G.M.;
 RT "Physical mapping and genomic structure of the human TNFR2 gene.";
 RL Genomics 35:94-100(1996).
 RN [4]
 RP SEQUENCE OF 37-461 FROM N.A.
 RX MEDLINE=91370690; PubMed=1966549;
 RA Dembic Z., Loetscher H., Gubler U., Pan Y.C., Lahm H.-W., Gentz R.,
 RA Brockhaus M., Lesslauer W.;
 RT "Two human TNF receptors have similar extracellular, but distinct
 RT intracellular, domain sequences.";
 RL Cytokine 2:231-237(1990).
 RN [5]
 RP SEQUENCE OF 116-461 FROM N.A., PARTIAL SEQUENCE, AND VARIANT ARG-196.
 RX MEDLINE=90349572; PubMed=2166946;
 RA Heller R.A., Song K., Onasch M.A., Fischer W.H., Chang D.,
 RA Ringold G.M.;
 RT "Complementary DNA cloning of a receptor for tumor necrosis factor
 RT and demonstration of a shed form of the receptor.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:6151-6155(1990).
 RN [6]
 RP SEQUENCE OF 154-183 FROM N.A., AND VARIANTS ARG-196 AND LYS-232.
 RX MEDLINE=21069356; PubMed=11197692;
 RA Tsuchiya N., Komata T., Matsushita M., Ohashi J., Tokunaga K.;
 RT "New single nucleotide polymorphisms in the coding region of human
 RT TNFR2: association with systemic lupus erythematosus.";
 RL Genes Immun. 1:501-503(2000).
 RN [7]
 RP SEQUENCE OF 27-31.
 RX MEDLINE=90110215; PubMed=2153136;
 RA Engelmann H., Novick D., Wallach D.;
 RT "Two tumor necrosis factor-binding proteins purified from human
 RT urine. Evidence for immunological cross-reactivity with cell surface
 RT tumor necrosis factor receptors.";
 RL J. Biol. Chem. 265:1531-1536(1990).
 RN [8]
 RP SEQUENCE OF 23-40; 65-69; 136-141; 300-306 AND 346-362.
 RX MEDLINE=91056048; PubMed=2173696;
 RA Loetscher H., Schlaeger E.J., Lahm H.-W., Pan Y.-C.E., Lesslauer W.,
 RA Brockhaus M.;
 RT "Purification and partial amino acid sequence analysis of two
 RT distinct tumor necrosis factor receptors from HD60 cells.";
 RL J. Biol. Chem. 265:20131-20138(1990).
 RN [9]
 RP CHARACTERIZATION.
 RX MEDLINE=93016040; PubMed=1328224;
 RA Penhica D., Lam V.T., Mize N.K., Weber R.F., Lewis M., Fendly B.M.,
 RA Lipari M.T., Goeddel D.V.;
 RT "Biochemical properties of the 75-kDa tumor necrosis factor receptor.
 RT Characterization of ligand binding, internalization, and receptor
 RT phosphorylation.";
 RL J. Biol. Chem. 267:21172-21178(1992).
 RN [10]
 RP INTERACTION WITH TRAF2.
 RX MEDLINE=94349371; PubMed=8069916;
 RA Rothe M., Mong S.C., Henzel W.J., Goeddel D.V.;
 RT "A novel family of putative signal transducers associated with the
 RT cytoplasmic domain of the 75 kDa tumor necrosis factor receptor.";
 RL Cell 78:681-692(1994).
 RN [11]
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 419-428 IN COMPLEX WITH
 RP TRAF2.


```

SQ SEQUENCE 462 AA; 51959 MW; 773A5DE377DF53F6 CRC64;
Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 462;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 470 LIERLND 476
|||||
DB 153 LIERLND 159

RESULT 83
VS1K.BMYVF STANDARD; PRT; 467 AA.
ID VS1K.BMYVF
AC P09514;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DE 51 kDa protein (ORF 6).
OS Beet western yellows virus (isolate FL-1) (BMYV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;
OC Polerovirus.
OC NCBI_TaxID=12043;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=89057523; PubMed=3194229;
RA Veidt I., Lot H., Leisner M., Scheidecker D., Guillely H., Richards K.,
RA Jonard G.;
RT "Nucleotide sequence of beet western yellows virus RNA.";
RL Nucleic Acids Res. 16:9917-9932(1988).
CC -1- SIMILARITY: TO ORF 4 OF BARLEY YELLOW DWARF VIRUS AND ORF6 OF
CC POTATO LEAFROLL VIRUS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X13063; CAA31467.1; -
CC DR PIR; S01943; S01943.
CC DR InterPro; IPR001517; Luteo coat.
CC DR InterPro; IPR000893; Luteo ORF6.
CC DR InterPro; IPR002929; PLRV ORF5.
CC DR Pfam; PF01690; PLRV_ORF5_1.
CC DR PRINTS; PR00910; LVIRUSORF6.
CC SQ SEQUENCE 467 AA; 51386 MW; 418C4E85FC23783 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 467;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 349 PSPQPTP 355
|||||
DB 13 PSPQPTP 19

RESULT 84
VS1K.BMYVG STANDARD; PRT; 471 AA.
ID VS1K.BMYVG
AC P09515;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DE 51 kDa protein (ORF 6).
OS Beet western yellows virus (isolate GBI) (BMYV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;
OC Polerovirus.
OC NCBI_TaxID=12044;
RN (1)
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=89057523; PubMed=3194229;
RA Veidt I., Lot H., Leisner M., Scheidecker D., Guillely H., Richards K.,
RA Jonard G.;
RT "Nucleotide sequence of beet western yellows virus RNA.";
RL Nucleic Acids Res. 16:9917-9932(1988).
CC -1- SIMILARITY: TO ORF 4 OF BARLEY YELLOW DWARF VIRUS AND ORF6 OF
CC POTATO LEAFROLL VIRUS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X13062; CAA31461.1; -
CC DR PIR; S01937; S01937.
CC DR InterPro; IPR000893; Luteo ORF6.
CC DR InterPro; IPR002929; PLRV ORF5.
CC DR Pfam; PF01690; PLRV_ORF5_1.
CC DR PRINTS; PR00910; LVIRUSORF6.
CC SQ SEQUENCE 471 AA; 52108 MW; DBE7E48070597EB8 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 471;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 349 PSPQPTP 355
|||||
DB 13 PSPQPTP 19

RESULT 85
YHGF.NEIGO STANDARD; PRT; 476 AA.
ID YHGF.NEIGO
AC Q51062;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein in region E (Fragment).
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OC NCBI_TaxID=485;
RN (1)
RP SEQUENCE FROM N.A.
RX STRAIN=MS11-E1;
RX MEDLINE=96236055; PubMed=8655518;
RA Petering H., Hammerschmidt S., Froesch M., van Putten J.P.M.,
RA Ison C.A., Robertson B.D.;
RT "Genes associated with meningococcal capsule complex are also found
RT in Neisseria gonorrhoeae.";
RL J. Bacteriol. 178:3342-3345(1996).
CC -1- SIMILARITY: Contains 1 SI motif domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z21508; CAA79717.1; -
CC DR InterPro; IPR006441; YggFc.
CC DR SMART; SM00732; YggFc_1.
CC KW Hypothetical protein; RNA-binding.
CC FT NON_TER 476 476
CC SQ SEQUENCE 476 AA; 52830 MW; 389E22AC91B8E065 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 476;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 201 SASELAA 207
Db 413 SASELAA 419

RESULT 86
VGLC_HSV2G STANDARD; PRT; 479 AA.

AC P03173;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Glycoprotein C precursor (Glycoprotein F).
GN GC OR UL44.
OS Herpes simplex virus (type 2 / strain G).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxId=10314;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85009847; PubMed=6090692;
RA Dombenko D.J., Lasky L.A.;
RT "Extensive homology between the herpes simplex virus type 2 glycoprotein F gene and the herpes simplex virus type 1 glycoprotein C gene";
RL J. Virol. 52:154-163(1984).
CC -1- MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1 AND 2: GH, GB, GC, GD, GI, AND GE.
CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN C FAMILY.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X01456; CAA25687.1; -.
CC PIR; A03734; VGBEF2.
DR InterPro: IPR001038; GP13_EHV.
DR InterPro: IPR007110; IG-like.
DR InterPro: IPR001654; Marek_A.
DR Pfam; PF02124; Marek_A; 1.
DR PRINTS; PR00668; GLYCOPROTEINC.
DR PROSITE; PS50835; IG LIKE; 1.
KW Immunoglobulin domain; Glycoprotein; Transmembrane; Signal.
FT SIGNAL 1 27
FT CHAIN 1 27
FT DOMAIN 25 327
FT CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 51 51 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 116 116 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 149 149 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 479 AA; 51667 MW; 9EDF4075108A933 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 479;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 290 VFDPQOI 296
Db 276 VFDPQOI 282

RESULT 87
VGLC_HSV23 STANDARD; PRT; 480 AA.

AC P06475;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Glycoprotein C precursor.
GN GC OR UL44.
OS Herpes simplex virus (type 2 / strain 333).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxId=10313;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85108149; PubMed=2982036;
RA Swain M.A., Peet R.W., Galloway D.A.;
RT "Characterization of the gene encoding herpes simplex virus type 2 glycoprotein C and comparison with the type 1 counterpart";
RL J. Virol. 53:561-569(1985).
CC -1- MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1 AND 2: GH, GB, GC, GD, GI, AND GE.
CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN C FAMILY.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U12178; AAA20532.1; -.
CC EMBL; X01996; CAA26025.1; -.
CC EMBL; M10053; AAA66442.1; -.
DR InterPro: IPR001038; GP13_EHV.
DR InterPro: IPR007110; IG-like.
DR InterPro: IPR001654; Marek_A.
DR Pfam; PF02124; Marek_A; 1.
DR PRINTS; PR00668; GLYCOPROTEINC.
DR PROSITE; PS50835; IG LIKE; 1.
KW Immunoglobulin domain; Glycoprotein; Transmembrane; Signal.
FT SIGNAL 1 27
FT CHAIN 1 27
FT DOMAIN 28 480
FT CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 52 52 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 117 117 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 150 150 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 480 AA; 51611 MW; CBB0F5BFE42A85 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 480;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 290 VFDPQOI 296
Db 277 VFDPQOI 283

RESULT 88
VGLC_HSV2H STANDARD; PRT; 480 AA.

AC Q89730; O12512;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Glycoprotein C precursor.
GN GC OR UL44.
OS Herpes simplex virus (type 2 / strain HG52).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxId=10315;

RN [1]
 RA SEQUENCE FROM N.A.
 RA Dolan A.;
 RN Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RC STRAIN=WTW1A;
 RA Terhune S.S., Spear P.G.;
 RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1 AND
 CC 2: GH, GB, GC, GG, GP, GI, AND GE.
 CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN C FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z86099; CAB06730.1; -
 DR EMBL; U12179; AAB60551.1; -
 DR EMBL; U12177; AAB60550.1; -
 DR InterPro: IPR001038; GP13 EHv.
 DR InterPro: IPR007110; IG-like.
 DR InterPro: IPR001654; Marek_A.
 DR Pfam: PF02124; Marek_A.1.
 DR PRINTS; PR00668; GLYCOPROTEIN.
 DR PROSITE; PS50835; IG_LIKE.1.
 KM Immunoglobulin domain; Glycoprotein; Transmembrane; Signal.
 FT SIGNAL 1 27
 FT CHAIN 28 480
 FT DOMAIN 236 328
 FT CARBOHYD 40 40
 FT CARBOHYD 52 52
 FT CARBOHYD 117 117
 FT CARBOHYD 150 150
 FT CARBOHYD 166 166
 FT CARBOHYD 331 331
 FT CARBOHYD 480 AA; 51671 MW; C145B4EFEF82B63 CRC64;
 SQ SEQUENCE
 Query Match 0.9%; Score 7; DB 1; Length 480;
 Best Local Similarity 100.0%; Pred No.1.le+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RP SEQUENCE FROM N.A.
 RA Coyne K.E., Johnson J.L., Rajagopalan K.V.;
 RT "Genomic DNA sequence of human sulfite oxidase SUOX."
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP VARIANTS GLN-160; ASP-208; TYR-370 AND ASP-473.
 RX MEDLINE=98088796; PubMed=9428520;
 RA Kisker C., Schindelin H., Pacheco A., Wehli W.A., Garrett R.M.,
 RA Rajagopalan K.V., Enemark J.H., Rees D.C.;
 RT "Molecular basis of sulfite oxidase deficiency from the structure of
 RT sulfite oxidase."
 RL Cell 91:973-983(1997).
 RN [4]
 RP VARIANT GLN-160.
 RX MEDLINE=98263367; PubMed=9600976;
 RA Garrett R.M., Johnson J.L., Graf T.N., Feigenbaum A.,
 RA Rajagopalan K.V.;
 RT "Human sulfite oxidase R160G: identification of the mutation in a
 RT sulfite oxidase-deficient patient and expression and characterization
 RT of the mutant enzyme."
 RL Proc. Natl. Acad. Sci. U.S.A. 95:6394-6398(1998).
 CC -1- CATALYTIC ACTIVITY: Sulfite + O(2) + H(2)O = sulfate + H(2)O(2).
 CC -1- COFACTOR: MOLYBDENUM (MOLYBDOPTERIN) AND ONE PROTHIEME GROUP.
 CC -1- PATHWAY: TERMINAL REACTION IN THE OXIDATIVE DEGRADATION OF SULFUR-
 CC CONTAINING AMINO ACIDS. IT USES CYTOCHROME C AS AN ELECTRON
 CC ACCEPTOR.
 CC -1- SUBUNIT: Homodimer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Mitochondrial intermembrane space.
 CC -1- DISEASE: DEFECTS IN SUOX ARE A CAUSE OF SULFITE OXIDASE
 CC DEFICIENCY, CHARACTERIZED BY NEUROLOGICAL ABNORMALITIES. OFTEN
 CC LEADS TO DEATH AT AN EARLY AGE.
 CC -1- SIMILARITY: WITH CYTOCHROME B5 AND NITRATE REDUCTASE.
 CC -1- SIMILARITY: CONTAINS 1 CYTOCHROME B5 HEME-BINDING DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L31573; AAA74886.1; -
 DR EMBL; AY056018; AAL08048.1; -
 DR PIR; S55874; S55874.
 DR PDB; 1M04; 12-SEP-02.
 DR Genew; HGNC:11460; SUOX.
 DR MIM; 606887; -
 DR MIM; 272300; -
 DR GO; GO:0008482; F:sulfite oxidase activity; TAS.
 DR InterPro: IPR001199; Cyt_B5.
 DR InterPro: IPR000572; Euk_Mb_oxred.
 DR InterPro: IPR005066; Mo-co_dimer.
 DR Pfam; PF00173; heme_1; 1.
 DR Pfam; PF03404; Mo-co_dimer; 1.
 DR Pfam; PF00174; oxidored_mol_yb; 1.
 DR PRINTS; PR00363; CYTOCHROME B5.
 DR PRINTS; PR00407; EUMOPTERIN.
 DR PRODOM; PD000612; Cyt_B5; 1.
 DR PROSITE; PS00191; CYTOCHROME B5_1; 1.
 DR PROSITE; PS50255; CYTOCHROME B5_2; 1.
 DR PROSITE; PS00559; MOLYBDOPTERIN_EUK; 1.
 KM Oxidoreductase; Mitochondrion; Heme; Molybdenum; Transit peptide;
 KW Disease mutation; 3D-structure.
 FT TRANSIT 1 22
 FT CHAIN 23 488
 FT DOMAIN 23 107
 FT DOMAIN 108 124
 FT DOMAIN 125 488
 FT METAL 61 61
 FT METAL 86 86
 FT METAL 207 207
 FT METAL 260 260
 HINGE (BY SIMILARITY).
 MOLYBDENUM-PTERIN DOMAIN (BY SIMILARITY).
 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
 MOLYBDENUM-PTERIN (BY SIMILARITY).
 MOLYBDENUM-PTERIN (BY SIMILARITY).

```

FT VARIANT 160 160 R -> Q (IN SUOX DEFICIENCY; 2% OF
FT ACTIVITY).
FT /FTID=VAR_002200.
FT VARIANT 208 208 A -> D (IN SUOX DEFICIENCY).
FT /FTID=VAR_002201.
FT VARIANT 370 370 S -> Y (IN SUOX DEFICIENCY).
FT /FTID=VAR_002202.
FT VARIANT 473 473 G -> D (IN SUOX DEFICIENCY).
FT /FTID=VAR_002203.
SQ SEQUENCE 488 AA, 53884 MW, 41EPA367FAB766DA CRC64;

Query Match 0.9%; Score 7; DB 1; Length 488;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 733 EAEVLLA 739
DB 286 EAEVLLA 292

RESULT 90
K2M2_SHEEP STANDARD; PRT; 491 AA.
ID K2M2_SHEEP
AC P15241;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Keratin, type II microfibrillar, component 7C.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE.
RA MEDLINE=90026244; PubMed=2803221;
RA Sparrow L.G., Robinson C.P., McMahon D.T.W., Rubira M.R.;
RT "The amino acid sequence of component 7c, a type II intermediate-
RL filament protein from wool.";
RL Biochem. J. 261:1013-1022(1989).
CC CC -I- FUNCTION: WOOL MICROFIBRILLAR KERATIN.
CC -I- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND
CC MICROFIBRILLAR KERATIN, I AND II (40-55 AND 56-70 KILODALTONS,
CC RESPECTIVELY). AT LEAST ONE TYPE OF CYTOSKELETAL KERATIN IS
CC PRESENT IN ALL VERTEBRATE EPITHELIAL CELLS.
CC -I- MISCELLANEOUS: THE LOW-SULFUR PROTEINS, DERIVED FROM THE
CC MICROFIBRILLAR FRACTION OF WOOL EXTRACTS, ARE COMPOSED OF TWO
CC FAMILIES, EACH CONSISTING OF 4 HOMOLOGOUS SUBUNITS: THE TYPE I
CC COMPONENTS (8C-1, 8C-2, 8A & 8B) AND THE TYPE II COMPONENTS (5,
CC 7A, 7B, AND 7C).
CC -I- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
DR InterPro; IPR001664; IF.
DR InterPro; IPR003054; Keratin_II.
DR Pfam; PF00038; filament_1.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Keratin.
FT MOD_RES 1 109 BLOCKED.
FT DOMAIN 1 109 HEAD.
FT DOMAIN 110 416 ROD.
FT DOMAIN 417 491 TAIL.
FT DOMAIN 110 144 COIL_1A.
FT DOMAIN 145 154 COIL_1B.
FT DOMAIN 155 255 COIL_1B.
FT DOMAIN 256 272 LINKER_12.
FT DOMAIN 273 416 COIL_2.
FT VARIANT 74 74 C -> G OR S.
FT VARIANT 80 80 C -> S.
FT VARIANT 144 144 F -> Y.
FT VARIANT 232 232 S -> V.
FT VARIANT 276 276 C -> D OR N.
FT VARIANT 284 284 Q -> H.
FT UNSURE 1 2 CG -> GC.

```

```

SQ SEQUENCE 491 AA, 53681 MW, AB01771FE3831ABE CRC64;

Query Match 0.9%; Score 7; DB 1; Length 491;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 RTKEIN 133
DB 320 RTKEIN 326

RESULT 91
K2M3_SHEEP STANDARD; PRT; 502 AA.
ID K2M3_SHEEP
AC P26591;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Keratin, type II microfibrillar, component 5.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE.
RA MEDLINE=92171886; PubMed=1371668;
RA Sparrow L.G., Robinson C.P., Caine J., McMahon D.T.W., Strike P.M.;
RT "Type II intermediate-filament proteins from wool. The amino acid
RL sequence of component 5 and comparison with component 7c.";
RL Biochem. J. 282:291-297(1992).
CC CC -I- FUNCTION: WOOL MICROFIBRILLAR KERATIN.
CC -I- TISSUE SPECIFICITY: HARD KERATIN WOOL.
CC -I- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND
CC MICROFIBRILLAR KERATIN, I AND II (40-55 AND 56-70 KILODALTONS,
CC RESPECTIVELY). AT LEAST ONE TYPE OF CYTOSKELETAL KERATIN IS
CC PRESENT IN ALL VERTEBRATE EPITHELIAL CELLS.
CC -I- MISCELLANEOUS: THE LOW-SULFUR PROTEINS, DERIVED FROM THE
CC MICROFIBRILLAR FRACTION OF WOOL EXTRACTS, ARE COMPOSED OF TWO
CC FAMILIES, EACH CONSISTING OF 4 HOMOLOGOUS SUBUNITS: THE TYPE I
CC COMPONENTS (8C-1, 8C-2, 8A & 8B) AND THE TYPE II COMPONENTS (5,
CC 7A, 7B, AND 7C).
CC -I- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
DR InterPro; IPR001664; IF.
DR InterPro; IPR003054; Keratin_II.
DR Pfam; PF00038; filament_1.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Keratin.
FT MOD_RES 1 122 BLOCKED.
FT DOMAIN 1 122 HEAD.
FT DOMAIN 123 429 ROD.
FT DOMAIN 430 502 TAIL.
FT DOMAIN 123 157 COIL_1A.
FT DOMAIN 158 167 LINKER_1.
FT DOMAIN 168 268 COIL_1B.
FT DOMAIN 269 285 LINKER_12.
FT DOMAIN 286 429 COIL_2.
FT UNSURE 1 2 SC -> CS.
SQ SEQUENCE 502 AA, 55255 MW, 8734C823050CE68 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 502;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 RTKEIN 133
DB 333 RTKEIN 339

RESULT 92
V56K_PLRV1

```



```

ID V56K_PLRV1 STANDARD; PRT; 508 AA.
AC P17525;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-AUG-1990 (Rel. 15, Last annotation update)
DE 56 kDa protein (ORF 6).
OC Potato leafroll virus (strain 1) (PLRV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;
OC Polerovirus.
OC NCBI_TaxID=12046;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89279282; PubMed=2732710;
RA Mayo M.A., Robinson D.J., Jolly C.A., Hyman L.;
RT "Nucleotide sequence of potato leafroll luteovirus RNA.";
RL J. Gen. Virol. 70:1037-1051(1989).
CC -1- SIMILARITY: 59% SIMILARITY TO ORF 6 OF BEET WESTERN YELLOWS
CC VIRUS AND 45% SIMILARITY TO ORF4 OF BARLEY YELLOW DWARF VIRUS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D00530; -; NOT_ANNOTATED_CDS.
DR PIR; S24594; S24594.
DR InterPro; IPR000893; Luteo_ORF6.
DR InterPro; IPR002929; PLRV_ORF5.
DR Pfam; PF01690; PLRV_ORF5; 1.
DR PRINTS; PR00910; LVIRUSORF6.
SQ SEQUENCE 508 AA; 56593 MW; FCB68654EF1E3D7 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 508;
Best Local Similarity 100.0%; Pred. No. 1.le+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 349 PSPQPTP 355
DB 9 PSPQPTP 15

RESULT 93
V56K_PLRV1 STANDARD; PRT; 508 AA.
AC P11626;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-AUG-1990 (Rel. 15, Last annotation update)
DE 56 kDa protein (ORF 6).
OC Potato leafroll virus (strain Wageningen) (PLRV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;
OC Polerovirus.
OC NCBI_TaxID=12048;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89171329; PubMed=2466700;
RA van der Wilk F., Huismans M.J., Cornelissen B.J.C., Huttinga H.,
RA Goldbach R.W.;
RT "Nucleotide sequence and organization of potato leafroll virus
RT genomic RNA.";
RL FEBS Lett. 245:51-56(1989).
CC -1- SIMILARITY: 59% SIMILARITY TO ORF4 OF BARLEY YELLOW DWARF VIRUS.
CC VIRUS AND 45% SIMILARITY TO ORF4 OF BARLEY YELLOW DWARF VIRUS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Y07496; CAA68799.1; ALT_INIT.
CC PIR; S03551; S03551.
CC InterPro; IPR000893; Luteo_ORF6.
CC InterPro; IPR002929; PLRV_ORF5.
CC Pfam; PF01690; PLRV_ORF5; 1.
CC PRINTS; PR00910; LVIRUSORF6.
SQ SEQUENCE 508 AA; 56510 MW; 195A346E9749C8A8 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 508;
Best Local Similarity 100.0%; Pred. No. 1.le+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 349 PSPQPTP 355
DB 9 PSPQPTP 15

RESULT 94
GUA_CLOPE STANDARD; PRT; 509 AA.
AC O8X146;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE GMP synthase [glutamine-hydrolyzing] (EC 6.3.5.2) (Glutamine
DE amidotransferase) (GMP synthetase).
GN GUA OR CPE2275.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
ON NCBI_TaxID=1502;
RX [1]
RP SEQUENCE FROM N.A.
RX STRAIN=13 / Type A;
RX MEDLINE=21664373; PubMed=11792842;
RA Shimizu T., Ohnami K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
CC -1- CATALYTIC ACTIVITY: ATP + xanthosine 5'-phosphate + L-glutamine +
CC H(2)O = AMP + diphosphate + GMP + L-glutamate.
CC -1- PATHWAY: GMP biosynthesis.
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE GMP SYNTHASE
CC FAMILY.
CC -1- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF003193; BAB81981.1; -.
DR HAMAP; MF_00344; -; 1.
DR InterPro; IPR006220; Anch synthet.
DR InterPro; IPR001317; CPS GATase.
DR InterPro; IPR000991; GATase_1.
DR InterPro; IPR001674; GMP_synth_C.
DR InterPro; IPR004739; GMPsynthase_N.
DR Pfam; PF00117; GATase; 1.
DR Pfam; PF00958; GMP synth C; 1.
DR PRINTS; PR00097; ANTSYNTHASEII.
DR PRINTS; PR00099; CPSGATASE.
DR PRINTS; PR00096; GATASE.
DR TIGRFAMs; TIGR00884; guaA_Cterm; 1.
DR TIGRFAMs; TIGR00888; guaA_Nterm; 1.
DR PROSITE; PS00442; GATASE_TYPE_1; 1.

```

KM Ligase; GMP biosynthesis; Purine biosynthesis; ATP-binding;
 KM Glutamine amidotransferase; Complete proteome.
 FT DOMAIN 1 193 GLUTAMINE AMIDOTRANSFERASE.
 FT NP_BIND 222 228 ATP (BY SIMILARITY).
 FT DOMAIN 226 385 GMP-BINDING (BY SIMILARITY).
 FT ACT_SITE 81 81 GATASE (BY SIMILARITY).
 FT ACT_SITE 168 168 GATASE (BY SIMILARITY).
 FT ACT_SITE 170 170 GATASE (BY SIMILARITY).
 SQ SEQUENCE 509 AA; 56843 MW; BAA75E1207DB029A CRC64;

 Query Match 0.9%; Score 7; DB 1; Length 509;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 97 IVEVKG 103
 DB 484 IVEVKG 490

 RESULT 95
 ID GUAA_CLOAB STANDARD; PRT; 510 AA.
 AC Q97FW9;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE GMP synthase [glutamine-hydrolyzing] (EC 6.3.5.2) (Glutamine
 amidotransferase) (GMP synthetase).
 GN GUAA OR CAC2700.
 OS Clostridium acetobutylicum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 CX NCBI_TaxID=1488;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
 RX MEDLINE=21359325; PubMed=1146286;
 RA Noelling J., Breton G., Omechenko M.V., Makarova K.S., Zeng Q.,
 RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
 RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
 RA Bennett G.N., Koonin E.V., Smith D.R.;
 RT "Genome sequence and comparative analysis of the solvent-producing
 bacterium Clostridium acetobutylicum";
 RL J. Bacteriol. 183:4823-4838(2001).
 CC -1- CATALYTIC ACTIVITY: ATP + xanthosine 5'-phosphate + L-glutamine +
 H(2)O = AMP + diphosphate + GMP + L-glutamate.
 CC -1- PATHWAY: GMP biosynthesis.
 CC -1- SUBUNIT: Homodimer (By similarity).
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE GMP SYNTHASE
 FAMILY.
 CC -1- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE007767; AAK80646.1; -
 CC PIR: C97232; C97232.
 DR HAMAP, MF_00344; -; 1.
 DR InterPro: IPR006220; Anth_synthII.
 DR InterPro: IPR001317; CPS_GATase.
 DR InterPro: IPR000991; GATase_1.
 DR InterPro: IPR001674; GMP_synth_C.
 DR InterPro: IPR004739; GMPsynthase_N.
 DR InterPro: IPR004506; TrnM.
 DR Pfam: PF00117; GATase_1.
 DR Pfam: PF00958; GMP_synth_C_1.
 DR Pfam: PF03054; rRNA_Me_trans_1.
 DR PRINTS; PR00097; ANTSNTHASEII.

DR PRINTS; PR00099; CPBGATASE.
 DR PRINTS; PR00096; GATASE.
 DR TIGRFAMs; TIGR00884; guaA_Cterm; 1.
 DR TIGRFAMs; TIGR00888; guaA_Nterm; 1.
 DR PROSITE; PS00442; GATASE_TYPE_1; 1.
 KM Ligase; GMP biosynthesis; Purine biosynthesis; ATP-binding;
 KM Glutamine amidotransferase; Complete proteome.
 FT DOMAIN 1 194
 FT NP_BIND 223 229 ATP (BY SIMILARITY).
 FT DOMAIN 227 386 GMP-BINDING (BY SIMILARITY).
 FT ACT_SITE 82 82 GATASE (BY SIMILARITY).
 FT ACT_SITE 169 169 GATASE (BY SIMILARITY).
 FT ACT_SITE 171 171 GATASE (BY SIMILARITY).
 SQ SEQUENCE 510 AA; 57269 MW; 066C9B02F850CAC CRC64;

 Query Match 0.9%; Score 7; DB 1; Length 510;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 97 IVEVKG 103
 DB 485 IVEVKG 491

 RESULT 96
 ID LYN_HUMAN STANDARD; PRT; 511 AA.
 AC P07948;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Tyrosine-protein kinase LYN (EC 2.7.1.112).
 GN LYN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87172710; PubMed=3561390;
 RA Yananashi Y., Fukushige S.-I., Semba K., Sukegawa J., Miyajima N.,
 RA Matsubara K.-I., Yamamoto T., Toyoshima K.;
 RT "The yes-related cellular gene lyn encodes a possible tyrosine kinase
 similar to p56lck";
 RL Mol. Cell. Biol. 7:237-243(1987).
 CC [2]
 CC SEQUENCE FROM N.A.
 RX MEDLINE=94171041; PubMed=8125304;
 RA Rider L.G., Raben N., Miller L., Jelsma C.;
 RT "The cdnas encoding two forms of the lyn protein tyrosine kinase are
 RT expressed in rat mast cells and human myeloid cells";
 RL Gene 138:219-222(1994).
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=LYN A;
 CC IsoId=P07948-1; Sequence=displayed;
 CC Name=LYN B;
 CC IsoId=P07948-2; Sequence=VSP_005002;
 CC -1- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. SRC
 CC SUBFAMILY.
 CC -1- SIMILARITY: Contains 1 SH3 domain.
 CC -1- SIMILARITY: Contains 1 SH3 domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: M16038; AAA59540.1; --
 DR EMBL: M79321; AAB50019.1; --
 DR PIR: A26719; TVHULY.
 DR HSSP: P08631; IAD5.
 DR Gene: HGNC:6735; LYN.
 DR MIM: 165120; --
 DR GO: GO:0004716; F:receptor signaling protein tyrosine kinase . . .; TAS.
 DR GO: GO:0006468; P:protein amino acid phosphorylation; TAS.
 DR InterPro: IPR000719; Prot_kinase.
 DR InterPro: IPR000980; SH2.
 DR InterPro: IPR001452; SH3.
 DR InterPro: IPR001245; Tyr_kinase.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF00017; SH2; 1.
 DR Pfam: PF00018; SH3; 1.
 DR PRINTS: PR00401; SH2DOMAIN.
 DR PRINTS: PR00452; SH3DOMAIN.
 DR PRINTS: PR00108; TYRKINASE.
 DR ProDom: PD000001; Prot_kinase; 1.
 DR ProDom: PD000093; SH2; 1.
 DR ProDom: PD000066; SH3; 1.
 DR SMART: SM00252; SH2; 1.
 DR SMART: SM00326; SH3; 1.
 DR SMART: SM00219; TYRC; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS50002; SH3; 1.
 DR PROSITE: PS50002; SH3; 1.
 DR PROTO-oncogene; Tyrosine-protein kinase; Phosphorylation;
 KW Transferrase; ATP-binding; Myristate; SH2 domain; SH3 domain;
 KW Palmitate; Lipoprotein; Alternative splicing.
 FT INIT MET 0
 FT LIPID 1 1 MYRISTATE (BY SIMILARITY).
 FT LIPID 2 2 PALMITATE (BY SIMILARITY).
 FT DOMAIN 62 122 SH2.
 FT DOMAIN 128 225 SH2.
 FT DOMAIN 246 500 PROTEIN_KINASE.
 FT NP_BIND 252 260 ATP (BY SIMILARITY).
 FT BINDING 274 274 ATP (BY SIMILARITY).
 FT ACT_SITE 366 366 BY SIMILARITY.
 FT MOD_RES 366 366 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 507 507 PHOSPHORYLATION (BY SIMILARITY).
 FT MOD_RES 507 507 Missing (in isoform LYN B).
 FT VARSPLIC 22 42 /FTId=VSP_005002.
 SO SEQUENCE 511 AA; 58442 MW; 8419CD461204B364 CRC64;
 Query Match 0.9%; Score 7; DB 1; Length 511;
 Best Local Similarity 100.0%; Pred. No. 1; le+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 557 GKDSLSD 563
 DB 7 GKDSLSD 13
 RESULT 97
 ID C84A_ARATH STANDARD; PRT; 520 AA.
 AC Q42600;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cytochrome P450 84A1 (Ferulate-5-hydroxylase) (EC 1.14.-.-) (F5H).
 GN CYP84A1 OR FAH1 OR AT4G36220 OR F2J313.110.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucoideae II; Brassicales; Brassicaceae; Arabidopsids.
 OC eucoideae II; Brassicales; Brassicaceae; Arabidopsids.
 OK NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;

RX MEDLINE=96293440; PubMed=8692910;
 RA Meyer K., Cusumano J.C., Somerville C.R., Chapple C.C.S.;
 RT "Ferulate-5-hydroxylase from Arabidopsis thaliana defines a new
 RL family of cytochrome P450-dependent monooxygenases.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:6869-6874(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Landsberg erecta;
 RX MEDLINE=99097044; PubMed=9880351;
 RA Ruegger M., Meyer K., Cusumano J.C., Chapple C.;
 RT "The regulation of ferulate-5-hydroxylase expression in Arabidopsis in
 RL the context of sinapate ester biosynthesis.";
 RL Plant Physiol. 119:101-110(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Many strains;
 RX MEDLINE=20580555; PubMed=11141187;
 RA Aguade M.;
 RT "Nucleotide sequence variation at two genes of the phenylpropanoid
 RL pathway, the FAH1 and F3H genes, in Arabidopsis thaliana.";
 RL Mol. Biol. Evol. 18:1-9(2001).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20083488; PubMed=10617198;
 RA Mayer K.F.X., Scheller C., Wambutt R., Murphy G., Voelckers G.,
 RA Pohl T., Duesterhoeft A., Striekema W., Ertler K.-D., Terryn N.,
 RA Harris B., Anesorge W., Brandt P., Grivell L.A., Rieger M.,
 RA Weichselgartner M., de Simone V., Obermayer B., Maché R., Mueller M.,
 RA Kreis M., Delsen M., Pilgomech P., Watson M., Schmidtke T.,
 RA Reichert B., Portelle D., Perez-Alonso M., Boutry M., Bancroft I.,
 RA Vos P., Hobeisel J., Zimmermann W., Wedler H., Ridley P.,
 RA Langham S.-A., McCallagh B., Bilham L., Robben J.,
 RA van der Schueren J., Grymoprez B., Chuang Y.-T., Vandenbusche F.,
 RA Braeken M., Welljens I., Voet M., Baetiaens I., Aert R., Defoor E.,
 RA Weitzenegeger T., Bothé G., Ramsperger U., Hilbert H., Braun M.,
 RA Holzner E., Brandt A., Peters S., van Staveren M., Dirkse W.,
 RA Moeljaner S., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
 RA Bernieris A., Hempel L., Feldpausch M., Lamberth S., Van den Daele H.,
 RA De Keyser A., Buysheert C., Gielen J., Villarroel R., De Clercq R.,
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
 RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McAlay K., Mayes R.,
 RA Petek A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
 RA Borkova D., Bloecher H., Scharfe M., Grimm M., Loehner T.-H.,
 RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
 RA Cabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
 RA Neumann S., Argitrou A., Vitale D., Liguori R., Piravandi E.,
 RA Massenot O., Quigley F., Clabaud G., Mendenhall A., Feilber R.,
 RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
 RA Cheifor F., Cooke R., Berger C., Montfort A., Casacuberta E.,
 RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,
 RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
 RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bleike C.,
 RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
 RA Parnell L., Dedhia N., Gnoj L., Schurz K., Huang E., Spiegel L.,
 RA Sekhon M., Murray J., Sheel P., Cordes M., Abu-Threideh U.,
 RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
 RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
 RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
 RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
 RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
 RA Antoniou B., Zidanic M., Strong C., Sun H., Lamer B., Yordan C.,
 RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,
 RA Granat S., Shobdy N., Hasegawa A., Hamed A., Lodhi M., Johnson A.,
 RA Chen E., Maria M., Martienssen R., McCombie W.R.;
 RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
 RL thaliana.";
 RL Nature 402:769-777(1999).
 CC -!- PATHWAY: Phenylpropanoid biosynthesis.
 CC -!- SIMILARITY: Belongs to the cytochrome P450 family.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U38416; AAC9389.1; -
DR EMBL; AL161589; CAB80293.1; -
DR EMBL; AJ295566; CAC26922.1; -
DR EMBL; AJ295567; CAC26923.1; -
DR EMBL; AJ295568; CAC26924.1; -
DR EMBL; AJ295569; CAC26925.1; -
DR EMBL; AJ295570; CAC26926.1; -
DR EMBL; AJ295571; CAC26927.1; -
DR EMBL; AJ295572; CAC26928.1; -
DR EMBL; AJ295573; CAC26929.1; -
DR EMBL; AJ295574; CAC26930.1; -
DR EMBL; AJ295575; CAC26931.1; -
DR EMBL; AJ295576; CAC26934.1; -
DR EMBL; AJ295579; CAC26935.1; -
DR EMBL; AL022141; CAAL18128.1; -
DR EMBL; AF068574; AAD11580.1; -
DR PIR; T04591; T04591.
DR HSSP; P14779; IJRP2.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450; 1.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
DR Oxidoreductase; Monooxygenase; Transmembrane; Heme; Multigene family.
FT TRANSMEM 12 32
FT METAL 458 458 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ SEQUENCE 520 AA; 58720 MW; E812779AF5BF01BC CRC64;

Query Match 0.9%; Score 7; DB 1; Length 520;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 485 VDDLLAF 491
DB 278 VDDLLAF 284

RESULT 98
CP11_CANFA STANDARD; PRT; 524 AA.
AC P56550;
ID 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cytochrome P450 1A1 (EC 1.14.14.1) (CYP1A1) (DAH1).
GN CYP1A1.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Beagle; TISSUE=Liver;
RX MEDLINE=91042464; PubMed=2122230;
RA Uchida T., Komori M., Kitada M., Kamataki T.;
RT "Isolation of cDNAs coding for three different forms of liver
RT microsomal cytochrome P-450 from polychlorinated biphenyl-treated
RT beagle-dogs."
RL Mol. Pharmacol. 38:644-651(1990).
CC -I- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
CC MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN
CC NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. THIS ISOLYME SEEMS
CC RESPONSIBLE FOR METABOLISM OF 2,2',4,4',5,5'-HEXACHLOROBIPHENYL.
CC -I- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC oxidized flavoprotein + H(2)O.

```

```

CC -----
CC -I- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC -I- INDUCTION: BY POLYCHLORINATED BIPHENYL (PCB) IN LIVER AND KIDNEY.
CC -I- SIMILARITY: Belongs to the cytochrome P450 family.
CC -I- CAUTION: IT IS UNCERTAIN WHETHER MET-1, MET-2 OR MET-4 IS THE
CC INITIATOR.
DR PIR; C37222; C37222.
DR HSSP; P00179; IJRP2.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
DR Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
DR Microsome; Endoplasmic reticulum.
FT METAL 461 461 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ SEQUENCE 524 AA; 59209 MW; 0BBD2C651C13BB40 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 524;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 732 KEAEVLL 738
DB 169 KEAEVLL 175

RESULT 99
GCR2_YEAST STANDARD; PRT; 534 AA.
ID GCR2_YEAST
AC Q01722;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Glycolytic genes transcriptional activator GCR2.
GN GCR2 OR YNL199C OR N1374.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A. PubMed=1508187;
RX MEDLINE=92375051;
RA Uemura H., Jigami Y.;
RT "Role of GCR2 in transcriptional activation of yeast glycolytic
RT genes."
RL Mol. Cell. Biol. 12:3834-3842(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE=95242839; PubMed=7725799;
RA Joniaux J.-L., Coster F., Purnelle B., Goffeau A.;
RT "A 21.7 kb DNA segment on the left arm of yeast chromosome XIV
RT carries WHI3, GCR2, SPX18, SPX19, an homologue to the heat shock gene
RT SSB1 and 8 new open reading frames of unknown function."
RL Yeast 10:1639-1645(1994).
CC -I- FUNCTION: REQUIRED FOR THE EXPRESSION OF MOST GLYCOLYTIC GENES.
CC GCR1 AND GCR2 PROBABLY FUNCTION AS A TRANSCRIPTIONAL ACTIVATION
CC COMPLEX, GCR1 PROVIDING THE SPECIFIC DNA-BINDING FUNCTION AND
CC GCR2 PROVIDING THE ACTIVATION FUNCTION.
CC -I- SUBUNIT: FORMS A GCR1/GCR2 COMPLEX.
CC -I- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -I- SIMILARITY: SOME, TO GCR1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D10104; BAA00985.1; -
DR EMBL; X78898; CAA55509.1; -
DR EMBL; Z71475; CAA96097.1; -

```

DR PIR; S31300; S31300.
 DR TRANSFAC; T03370; -.
 DR SGD; S0005143; GCR2.
 DR GO; GO:0016563; F:transcriptional activator activity; IDA.
 DR GO; GO:0045821; P:positive regulation of glycolysis; IMP.
 KM Transcription regulation; Activator; Nuclear protein.
 FT DOMAIN 281 288 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 255 283 ASN-RICH.
 SQ SEQUENCE 534 AA; 58062 MW; 958D4A9393255B1B CRC64;

Query Match
 Best Local Similarity 100.0%; Score 7; DB 1; Length 534;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 75_YDAIISE 81
 Db 486_YDAIISE 492

RESULT 100
 GLE1_YEAST
 ID GLE1_YEAST STANDARD; PRT; 538 AA.
 AC Q12315;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE RNA export factor GLE1.
 GN GLE1 OR BRR3 OR YDL207W OR D1049.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A., AND MUTAGENESIS.
 RX MEDLINE=96404361; PubMed=8848052;
 RA Murphy R., Wente S.R.;
 RT "An RNA-export mediator with an essential nuclear export signal.";
 RL Nature 383:357-360 (1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA STRAIN=S288c / FY1679;
 RA Bahr A., Moeller-Rieker S., Hankeln T., Kraemer C., Schmidt E.R.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: RNA EXPORT FACTOR. THE MECHANISM MAY INVOLVE THE DIRECT
 TARGETING OF RNA/RNP MOLECULES TO OR THROUGH THE NUCLEAR PORE
 COMPLEX BY MEANS OF R1P1 AND NUP100 INTERACTIONS.
 CC -!- SUBCELLULAR LOCATION: NUCLEAR. LOCALIZED PREDOMINANTLY AT THE
 NUCLEAR PORE COMPLEXES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U68475; AAC9444.1; -.
 DR EMBL; X99000; CAA67484.1; -.
 DR EMBL; Z74255; CAA98785.1; -.
 DR PIR; S67766; S67766.
 DR SGD; S0002366; GLE1.
 DR GO; GO:0016973; P:poly(A)+ mRNA-nucleus export; IMP.
 KM Nuclear protein; mRNA processing.
 FT DOMAIN 351 358 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT MUTAGEN 351 351 L->A: PARTIAL LOSS OF ACTIVITY.
 FT MUTAGEN 353 353 L->A: PARTIAL LOSS OF ACTIVITY.
 FT MUTAGEN 356 356 L->A: TEMPERATURE SENSITIVE.
 FT MUTAGEN 358 358 L->A: PARTIAL LOSS OF ACTIVITY.
 SQ SEQUENCE 538 AA; 62073 MW; C9C6B1513AP95711 CRC64;

Query Match
 Best Local Similarity 100.0%; Score 7; DB 1; Length 538;
 Pred. No. 1.2e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 353_PTPESP 359
 Db 33_PTPESP 39

Search completed: November 14, 2003, 10:59:21
 Job time : 35 secs

THIS PAGE BLANK (USPTO)

**THIS PAGE BLANK (USPTO)
BEST AVAILABLE COPY**

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 10:45:15 ; Search time 88 Seconds
(without alignments)
2334.203 Million cell updates/sec

Title: US-09-765-271-56

Perfect score: 796

Sequence: 1 SYELGLYQARTVKENNRVSY.....KLIALIKGSPSSVSKKIN 796

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 200 summaries

Database :

SPTRMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeophages:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	795	99.9	816	2 Q9AHT9	Q9AHT9 streptococc
2	695	87.3	802	16 Q97QM8	Q97QM8 streptococc
3	616	77.4	828	16 Q8DPQ2	Q8DPQ2 streptococc
4	107	13.4	844	2 Q9AG74	Q9AG74 streptococc
5	107	13.4	855	16 Q8CWR4	Q8CWR4 streptococc
6	60	7.5	819	2 Q9ANY3	Q9ANY3 streptococc
7	60	7.5	819	16 Q97QM9	Q97QM9 streptococc
8	60	7.5	839	16 Q9ANY2	Q9ANY2 streptococc
9	60	7.5	853	16 Q8DQ08	Q8DQ08 streptococc
10	32	4.0	1039	16 Q9ANY1	Q9ANY1 streptococc
11	32	4.0	1039	16 Q8DQ07	Q8DQ07 streptococc
12	27	3.4	481	16 Q8E338	Q8E338 streptococc
13	17	2.1	289	2 Q9AE21	Q9AE21 streptococc
14	17	2.1	822	2 Q9ZHG7	Q9ZHG7 streptococc
15	17	2.1	822	16 Q8E4U1	Q8E4U1 streptococc
16	17	2.1	822	16 Q8D281	Q8D281 streptococc

17	17	2.1	823	16 Q8N282	Q8N282 streptococc
18	17	2.1	823	16 Q8K5Q1	Q8K5Q1 streptococc
19	17	2.1	825	2 Q93GTS	Q93GTS streptococc
20	17	2.1	825	16 Q99XV4	Q99XV4 streptococc
21	10	1.3	182	16 Q8DQ06	Q8DQ06 streptococc
22	9	1.1	791	16 Q9CLO1	Q9CLO1 pasteurella
23	8	1.0	71	12 Q8B9M0	Q8B9M0 rachidiplusia
24	8	1.0	147	16 Q8EHD8	Q8EHD8 shewanella
25	8	1.0	152	5 Q8IIB9	Q8IIB9 plasmidium
26	8	1.0	154	5 P91742	P91742 hydra atten
27	8	1.0	181	5 Q44095	Q44095 drosophila
28	8	1.0	270	5 Q9NCF0	Q9NCF0 cryptocercu
29	8	1.0	284	5 Q22866	Q22866 caenorhabdi
30	8	1.0	284	5 Q22865	Q22865 caenorhabdi
31	8	1.0	284	5 Q27284	Q27284 caenorhabdi
32	8	1.0	294	5 Q46010	Q46010 caenorhabdi
33	8	1.0	301	5 Q95Q31	Q95Q31 caenorhabdi
34	8	1.0	319	16 Q8FLF7	Q8FLF7 corneobacte
35	8	1.0	436	3 Q9P8N1	Q9P8N1 coriolus ve
36	8	1.0	484	16 Q9KS51	Q9KS51 vibrio chol
37	8	1.0	485	10 Q80401	Q80401 oryza sativ
38	8	1.0	488	16 Q99TD8	Q99TD8 staphylococ
39	8	1.0	488	16 Q8NW40	Q8NW40 staphylococ
40	8	1.0	490	16 Q8CNW2	Q8CNW2 staphylococ
41	8	1.0	673	6 Q95LH9	Q95LH9 oryctolagus
42	8	1.0	840	16 Q9RSJ4	Q9RSJ4 deinococcus
43	8	1.0	945	16 Q26077	Q26077 helicobacte
44	8	1.0	1233	17 Q8TNX4	Q8TNX4 methanobact
45	8	1.0	1446	2 Q93HW0	Q93HW0 streptococc
46	8	1.0	1659	16 Q97S90	Q97S90 streptococc
47	8	1.0	1659	16 Q8C252	Q8C252 streptococc
48	8	1.0	2621	5 Q8SSW5	Q8SSW5 dictyostell
49	7	0.9	42	15 Q87936	Q87936 chimpanzee
50	7	0.9	42	15 Q87948	Q87948 chimpanzee
51	7	0.9	42	15 Q87938	Q87938 chimpanzee
52	7	0.9	42	15 Q88084	Q88084 chimpanzee
53	7	0.9	42	15 Q87952	Q87952 chimpanzee
54	7	0.9	42	15 Q87946	Q87946 chimpanzee
55	7	0.9	42	15 Q88080	Q88080 chimpanzee
56	7	0.9	43	15 Q87944	Q87944 chimpanzee
57	7	0.9	44	15 Q87942	Q87942 chimpanzee
58	7	0.9	49	15 Q87940	Q87940 chimpanzee
59	7	0.9	49	16 Q45596	Q45596 bacillus su
60	7	0.9	50	2 Q9AGM9	Q9AGM9 clostridium
61	7	0.9	51	16 Q8E4V0	Q8E4V0 escherichia
62	7	0.9	68	15 Q90VA0	Q90VA0 human immun
63	7	0.9	68	16 Q9A3B0	Q9A3B0 caulobacter
64	7	0.9	71	1 Q977L2	Q977L2 uncultured
65	7	0.9	71	17 Q96YD8	Q96YD8 sulfobolus
66	7	0.9	80	3 Q9HGR7	Q9HGR7 gilbertella
67	7	0.9	83	5 Q8MYJ9	Q8MYJ9 dictyostell
68	7	0.9	86	15 Q73850	Q73850 human immun
69	7	0.9	86	15 Q73851	Q73851 human immun
70	7	0.9	86	15 Q73852	Q73852 human immun
71	7	0.9	90	17 Q97W98	Q97W98 sulfobolus
72	7	0.9	93	16 Q9RUC5	Q9RUC5 deinococcus
73	7	0.9	94	16 Q9RY27	Q9RY27 deinococcus
74	7	0.9	96	16 Q8FAP6	Q8FAP6 escherichia
75	7	0.9	104	17 Q8U029	Q8U029 pyrococcus
76	7	0.9	108	15 P88557	P88557 human immun
77	7	0.9	109	2 Q8VR72	Q8VR72 escherichia
78	7	0.9	110	16 Q9A654	Q9A654 caulobacter
79	7	0.9	110	16 Q9CDT0	Q9CDT0 lactococcus
80	7	0.9	111	16 Q8FC10	Q8FC10 escherichia
81	7	0.9	115	16 Q8CMN4	Q8CMN4 staphylococ
82	7	0.9	116	5 Q8T427	Q8T427 plasmidium
83	7	0.9	117	10 Q8L494	Q8L494 oryza sativ
84	7	0.9	117	16 Q9YC21	Q9YC21 bruceella me
85	7	0.9	120	12 Q8GUR7	Q8GUR7 infectious
86	7	0.9	124	16 Q8FDX0	Q8FDX0 escherichia
87	7	0.9	126	2 Q9X6J3	Q9X6J3 bacillus st
88	7	0.9	128	16 Q8FWA2	Q8FWA2 bruceella su
89	7	0.9	128	16 Q8EMH8	Q8EMH8 oceanobacil

90	7	0.9	128	17	Q9YAF7	Q9YAF7 aeropyrum p
91	7	0.9	129	16	Q9XBD3	Q9XBD3 streptomyces
92	7	0.9	130	16	P74726	P74726 synechocyst
93	7	0.9	131	2	O69802	O69802 streptomyces
94	7	0.9	131	16	Q92P78	Q92P78 rhizobium m
95	7	0.9	131	16	O86762	O86762 streptomyces
96	7	0.9	134	12	Q9YL97	Q9YL97 human adeno
97	7	0.9	134	15	O90640	O90640 simian-huma
98	7	0.9	134	15	O8PN57	O8PN57 xanthomonas
99	7	0.9	136	2	O8V819	O8V819 shigella fl
100	7	0.9	136	5	Q17947	Q17947 caenorhabdi
101	7	0.9	136	16	Q9JMT0	Q9JMT0 escherichia
102	7	0.9	137	15	O8QDG8	O8QDG8 chimpanzee
103	7	0.9	137	15	O8QDG9	O8QDG9 chimpanzee
104	7	0.9	137	15	O8QDG3	O8QDG3 chimpanzee
105	7	0.9	137	15	O8QDH0	O8QDH0 chimpanzee
106	7	0.9	137	15	O8QDH5	O8QDH5 chimpanzee
107	7	0.9	137	15	O8QDH1	O8QDH1 chimpanzee
108	7	0.9	137	15	O8QDH3	O8QDH3 chimpanzee
109	7	0.9	137	15	O8QDH2	O8QDH2 chimpanzee
110	7	0.9	137	15	O8QDG5	O8QDG5 chimpanzee
111	7	0.9	137	15	O8QDG4	O8QDG4 chimpanzee
112	7	0.9	137	15	O8QDG6	O8QDG6 chimpanzee
113	7	0.9	139	16	O9PMF7	O9PMF7 campylobact
114	7	0.9	143	16	O98J01	O98J01 rhizobium 1
115	7	0.9	144	6	O9BH14	O9BH14 antillocapra
116	7	0.9	145	16	O8Y7M1	O8Y7M1 listeria mo
117	7	0.9	147	16	O9HZE1	O9HZE1 pseudomonas
118	7	0.9	148	16	O8X8A4	O8X8A4 escherichia
119	7	0.9	149	16	O8DL26	O8DL26 synechococc
120	7	0.9	151	2	O9AFN4	O9AFN4 shigella fl
121	7	0.9	152	5	O8N052	O8N052 drosophila
122	7	0.9	152	5	O8N054	O8N054 xanthomonas
123	7	0.9	152	16	O8P1G4	O8P1G4 xanthomonas
124	7	0.9	152	16	O8P749	O8P749 xanthomonas
125	7	0.9	153	10	O9LVZ0	O9LVZ0 arabidopsis
126	7	0.9	153	10	O8N041	O8N041 drosophila
127	7	0.9	154	5	O8N048	O8N048 drosophila
128	7	0.9	154	5	O8N051	O8N051 drosophila
129	7	0.9	155	5	O8N042	O8N042 drosophila
130	7	0.9	155	5	O8N039	O8N039 drosophila
131	7	0.9	155	5	O8N013	O8N013 drosophila
132	7	0.9	155	5	O8N040	O8N040 drosophila
133	7	0.9	155	5	O8N047	O8N047 drosophila
134	7	0.9	155	5	O8N053	O8N053 drosophila
135	7	0.9	156	5	O8N053	O8N053 drosophila
136	7	0.9	156	11	Q9JUF0	Q9JUF0 mus musculu
137	7	0.9	157	5	O9Y064	O9Y064 taenia soli
138	7	0.9	157	5	O8N043	O8N043 drosophila
139	7	0.9	157	5	O8N054	O8N054 drosophila
140	7	0.9	157	5	O8N049	O8N049 drosophila
141	7	0.9	157	5	O8N045	O8N045 drosophila
142	7	0.9	157	16	O8UBA0	O8UBA0 agrobacteri
143	7	0.9	157	16	O8U631	O8U631 agrobacteri
144	7	0.9	157	16	O8PM65	O8PM65 xanthomonas
145	7	0.9	158	10	O8GWR9	O8GWR9 arabidopsis
146	7	0.9	159	2	O86437	O86437 escherichia
147	7	0.9	160	2	O8GHN8	O8GHN8 pseudomonas
148	7	0.9	160	5	O8N046	O8N046 drosophila
149	7	0.9	161	5	O8N030	O8N030 drosophila
150	7	0.9	161	5	O8N015	O8N015 drosophila
151	7	0.9	161	5	O8N035	O8N035 drosophila
152	7	0.9	162	5	O8N020	O8N020 drosophila
153	7	0.9	163	5	O8N036	O8N036 drosophila
154	7	0.9	163	5	O8N017	O8N017 drosophila
155	7	0.9	164	5	O8N032	O8N032 drosophila
156	7	0.9	164	5	O8N033	O8N033 drosophila
157	7	0.9	164	5	O8N029	O8N029 drosophila
158	7	0.9	164	5	O8N023	O8N023 drosophila
159	7	0.9	164	5	O8N028	O8N028 drosophila
160	7	0.9	164	5	O8N027	O8N027 drosophila
161	7	0.9	165	5	O8N026	O8N026 drosophila
162	7	0.9	165	5	O8N026	O8N026 drosophila

163	7	0.9	165	5	O8N024	O8N024 drosophila
164	7	0.9	165	5	O8N019	O8N019 drosophila
165	7	0.9	165	5	O8N034	O8N034 drosophila
166	7	0.9	165	5	O8N016	O8N016 drosophila
167	7	0.9	165	5	O8N031	O8N031 drosophila
168	7	0.9	165	5	O8N022	O8N022 drosophila
169	7	0.9	165	16	O912Q4	O912Q4 pseudomonas
170	7	0.9	165	5	O8N037	O8N037 drosophila
171	7	0.9	166	5	O8N021	O8N021 drosophila
172	7	0.9	166	16	O8NLR4	O8NLR4 corynebacte
173	7	0.9	167	5	O8N057	O8N057 drosophila
174	7	0.9	167	5	O8N025	O8N025 drosophila
175	7	0.9	167	5	O8N044	O8N044 drosophila
176	7	0.9	167	5	O8N038	O8N038 drosophila
177	7	0.9	167	16	O8DBU1	O8DBU1 vibrio vuln
178	7	0.9	168	2	O8GHB0	O8GHB0 chlamydomo
179	7	0.9	168	5	O8N018	O8N018 drosophila
180	7	0.9	169	5	O8N059	O8N059 drosophila
181	7	0.9	169	5	O8N055	O8N055 drosophila
182	7	0.9	169	5	O8N060	O8N060 drosophila
183	7	0.9	169	5	O8N058	O8N058 drosophila
184	7	0.9	171	5	O8N056	O8N056 drosophila
185	7	0.9	177	16	O8RIJ7	O8RIJ7 fusobacteri
186	7	0.9	181	16	O51571	O51571 borrelia bu
187	7	0.9	184	4	O96RK1	O96RK1 homo sapien
188	7	0.9	184	16	O8P8K5	O8P8K5 xanthomonas
189	7	0.9	184	16	O8FS32	O8FS32 corynebacte
190	7	0.9	186	2	O49530	O49530 mycoplasma
191	7	0.9	186	10	O9LID5	O9LID5 arabidopsis
192	7	0.9	186	17	O96E27	O96E27 sulfolobus
193	7	0.9	187	2	O8LRF4	O8LRF4 oryza sativ
194	7	0.9	189	2	O8VVE6	O8VVE6 salmonella
195	7	0.9	189	10	O8H341	O8H341 oryza sativ
196	7	0.9	190	16	O8CK90	O8CK90 yersinia pe
197	7	0.9	192	10	O9FIT9	O9FIT9 arabidopsis
198	7	0.9	192	11	O9DA99	O9DA99 mus musculu
199	7	0.9	198	4	O9NPS2	O9NPS2 homo sapien
200	7	0.9	198	17	O8ZV93	O8ZV93 pyrobaculum

ALIGNMENTS

RESULT 1

ID Q9AHT9 PRELIMINARY; PRT; 816 AA.

AC Q9AHT9; 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

DE Streptococcus histidine triad A protein.

GN PHTA.

OS Streptococcus pneumoniae.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

CX NCBI_TaxID=1313;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=N4;

RX MEDLINE=21116976; PubMed=1179332;

RA Wisemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,

RA Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,

RA Gayle A., Brewah Y.A., Walsh W., Barron P., Lathigra R., Hanson M.,

RA Langerman S., Johnson S., Koenig S.,

RT "Use of a Whole Genome Approach To Identify Vaccine Molecules

RT Affording Protection Against Streptococcus pneumoniae Infection.";

RL Infect. Immun. 69:1593-1598(2001).

DR EMBL: AF291695; AAK19155.1; "

DR InterPro: IPR006270; Strep_his_triad.

DR TIGRFAWS; TIGR01363; strep_his_triad; 2

SQ SEQUENCE 816 AA; 91519 MW; 5359126A61ID27ED CRC64;

Query Match 99.9%; Score 795; DB 2; Length 816;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYELGLYOARTVKNRRVSYIDGKQATQKTEMLTPDEVSKREGINAEQIVIKITDGGYVT 60
DB 21 SYELGLYOARTVKNRRVSYIDGKQATQKTEMLTPDEVSKREGINAEQIVIKITDGGYVT 80
QY 61 SHGDHYYNGKVPYDAIISEELMKDPNYLKDEDIVNEVGKGYVIRKDGKYYVYLKDA 120
DB 81 SHGDHYYNGKVPYDAIISEELMKDPNYLKDEDIVNEVGKGYVIRKDGKYYVYLKDA 140
QY 121 AHADNVRKKEINRQKQESHQHEGGTPRNDGVALARSQGRYTTDGYIFNADIIEDT 180
DB 141 AHADNVRKKEINRQKQESHQHEGGTPRNDGVALARSQGRYTTDGYIFNADIIEDT 200
QY 181 GDAYIVPHGDHYHIFPKNELSASELAFAFLSGRGLNSRTRYRQNSDNTSRTWVPS 240
DB 201 GDAYIVPHGDHYHIFPKNELSASELAFAFLSGRGLNSRTRYRQNSDNTSRTWVPS 260
QY 241 VSNPGTTNTNTSNNSNTNSQASQNDIDSLKQLYKLPSQRHVESDGLVFPDAQITSRT 300
DB 261 VSNPGTTNTNTSNNSNTNSQASQNDIDSLKQLYKLPSQRHVESDGLVFPDAQITSRT 320
QY 301 ARGVAVPHGDHYHIFPYSQMSELEERIIARIIPLRYSNMHWPPDSRPEQSPQPTPEPSPG 360
DB 321 ARGVAVPHGDHYHIFPYSQMSELEERIIARIIPLRYSNMHWPPDSRPEQSPQPTPEPSPG 380
QY 361 POPAPMLKIDSNSLSVQLVRKVGEGYFEEKISRYVFAKDLPSFTVKNLESKLSKQES 420
DB 381 POPAPMLKIDSNSLSVQLVRKVGEGYFEEKISRYVFAKDLPSFTVKNLESKLSKQES 440
QY 421 VSHITLAKKENVAAPROEFYDKAYNLLTEAHKALFNKGRNSDFQALDKLELNDESTN 480
DB 441 VSHITLAKKENVAAPROEFYDKAYNLLTEAHKALFNKGRNSDFQALDKLELNDESTN 500
QY 481 KEKLVDDLAFAPITHPERLGRKPSQIEYTEDEVRIAQLADKYTTSDGYIFDEHDIISD 540
DB 501 KEKLVDDLAFAPITHPERLGRKPSQIEYTEDEVRIAQLADKYTTSDGYIFDEHDIISD 560
QY 541 EGDAYVTPHMGSHHWGKSLSDKEKVAQAATTKEGGILPPSPDADVKNPFGDSAAAY 600
DB 561 EGDAYVTPHMGSHHWGKSLSDKEKVAQAATTKEGGILPPSPDADVKNPFGDSAAAY 620
QY 601 NRKVGKRIPLVRLPMVHEHTEVKNGNLIIPKHCHYHIKFAWPDHHTYKAPNGYTLSD 660
DB 621 NRKVGKRIPLVRLPMVHEHTEVKNGNLIIPKHCHYHIKFAWPDHHTYKAPNGYTLSD 680
QY 661 LFATIKYYVEHPDERPHSNDGNGNASEHVLGKKDHSDEPNKPKADEEVEETPAPEVP 720
DB 681 LFATIKYYVEHPDERPHSNDGNGNASEHVLGKKDHSDEPNKPKADEEVEETPAPEVP 740
QY 721 QVTEKEVAQLKEAEVLLAKVYTSSLKAAATEFLAGLRNNLTLOIMDNNSIMAEKLLA 780
DB 741 QVTEKEVAQLKEAEVLLAKVYTSSLKAAATEFLAGLRNNLTLOIMDNNSIMAEKLLA 800
QY 781 LKGSNPPSSVSKKIN 796
DB 801 LKGSNPPSSVSKKIN 816

RESULT 2
Q970M8 PRELIMINARY; PRT; 802 AA.
AC Q970M8; 01-OCT-2001 (Tremblrel. 18, Created)
DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
DE 01-MAR-2003 (Tremblrel. 23, Last annotation update)
GN SP1175.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
NCBI_Taxid=1313;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettein H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Whitm M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Lotz B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae";
RL Science 293:498-506(2001).
DR EMBL: AE007418; AKK75284.1; -.
DR TIGR: SP1175; -.
DR InterPro: IPR006270; strep_his_triad.
DR TIGRFams: TIGR01363; strep_his_triad; 2.
KW Complete proteome.
SQ SEQUENCE 802 AA; 90080 MW; 4E5CB8364EBA1833 CRC64;

Query Match 87.3%; Score 695; DB 16; Length 802;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYELGLYOARTVKNRRVSYIDGKQATQKTEMLTPDEVSKREGINAEQIVIKITDGGYVT 60
DB 7 SYELGLYOARTVKNRRVSYIDGKQATQKTEMLTPDEVSKREGINAEQIVIKITDGGYVT 66
QY 61 SHGDHYYNGKVPYDAIISEELMKDPNYLKDEDIVNEVGKGYVIRKDGKYYVYLKDA 120
DB 67 SHGDHYYNGKVPYDAIISEELMKDPNYLKDEDIVNEVGKGYVIRKDGKYYVYLKDA 126
QY 121 AHADNVRKKEINRQKQESHQHEGGTPRNDGVALARSQGRYTTDGYIFNADIIEDT 180
DB 127 AHADNVRKKEINRQKQESHQHEGGTPRNDGVALARSQGRYTTDGYIFNADIIEDT 186
QY 181 GDAYIVPHGDHYHIFPKNELSASELAFAFLSGRGLNSRTRYRQNSDNTSRTWVPS 240
DB 187 GDAYIVPHGDHYHIFPKNELSASELAFAFLSGRGLNSRTRYRQNSDNTSRTWVPS 246
QY 241 VSNPGTTNTNTSNNSNTNSQASQNDIDSLKQLYKLPSQRHVESDGLVFPDAQITSRT 300
DB 247 VSNPGTTNTNTSNNSNTNSQASQNDIDSLKQLYKLPSQRHVESDGLVFPDAQITSRT 306
QY 301 ARGVAVPHGDHYHIFPYSQMSELEERIIARIIPLRYSNMHWPPDSRPEQSPQPTPEPSPG 360
DB 307 ARGVAVPHGDHYHIFPYSQMSELEERIIARIIPLRYSNMHWPPDSRPEQSPQPTPEPSPG 366
QY 361 POPAPMLKIDSNSLSVQLVRKVGEGYFEEKISRYVFAKDLPSFTVKNLESKLSKQES 420
DB 367 POPAPMLKIDSNSLSVQLVRKVGEGYFEEKISRYVFAKDLPSFTVKNLESKLSKQES 426
QY 421 VSHITLAKKENVAAPROEFYDKAYNLLTEAHKALFNKGRNSDFQALDKLELNDESTN 480
DB 427 VSHITLAKKENVAAPROEFYDKAYNLLTEAHKALFNKGRNSDFQALDKLELNDESTN 486
QY 481 KEKLVDDLAFAPITHPERLGRKPSQIEYTEDEVRIAQLADKYTTSDGYIFDEHDIISD 540
DB 487 KEKLVDDLAFAPITHPERLGRKPSQIEYTEDEVRIAQLADKYTTSDGYIFDEHDIISD 546
QY 541 EGDAYVTPHMGSHHWGKSLSDKEKVAQAATTKEGGILPPSPDADVKNPFGDSAAAY 600
DB 547 EGDAYVTPHMGSHHWGKSLSDKEKVAQAATTKEGGILPPSPDADVKNPFGDSAAAY 606
QY 601 NRKVGKRIPLVRLPMVHEHTEVKNGNLIIPKHCHYHIKFAWPDHHTYKAPNGYTLSD 660
DB 607 NRKVGKRIPLVRLPMVHEHTEVKNGNLIIPKHCHYHIKFAWPDHHTYKAPNGYTLSD 666
QY 661 LFATIKYYVEHPDERPHSNDGNGNASEHVLGKKDHSDEPNKPKADEEVEETPAPEVP 720
DB 667 LFATIKYYVEHPDERPHSNDGNGNASEHVLGKKDHSDEPNKPKADEEVEETPAPEVP 726

```

OY 721 OVEETEKEVAQLKEAEVLLAKVTDSLLKANATETTLAGLRNNLTLOIMDNNSIMAEAEKLLA 780
DB 727 OVEETEKEVAQLKEAEVLLAKVTDSLLKANATETTLAGLRNNLTLOIMDNNSIMAEAEKLLA 786
OY 781 LKGSNPSSVSKEKIN 796
DB 787 LKGSNPSSVSKEKIN 802

RESULT 3
Q8DPQ2 PRELIMINARY; PRT; 828 AA.
AC Q8DPQ2:
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DE Pneumococcal histidine triad protein A.
GN PHPA OR SPRI061.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E., Jr., Arnold J., Blaszczak L.C., Burgett S.,
RA Dehoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA Leblanc D.J., Lee L.N., Leikowitz E.J., Lu J., Matsushima P.,
RA McAhren S.M., McHenry M., Mcleaster C., Mundy C.W., Nices T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rokey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Kosteck P.R., Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
DR EMBL: AE008479; AAK9865.1; -.
KW Complete proteome.
SQ SEQUENCE 828 AA; 93015 MW; 12CCCF407B550CID CRC64;

Query Match 77.4%; Score 616; DB 16; Length 828;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 716; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 80 SEELMKDPNRYLKDEDIVNEVKGVIKVDGKYVYLKDAHANVTKKEINRQOE 139
DB 112 SEELMKDPNRYLKDEDIVNEVKGVIKVDGKYVYLKDAHANVTKKEINRQOE 171
OY 140 SOHREGCPRNDGAVALARSGRYTTDDGYIFNADIIEDTGDAYIVPHGDHYHIPNE 199
DB 172 SOHREGCPRNDGAVALARSGRYTTDDGYIFNADIIEDTGDAYIVPHGDHYHIPNE 231
OY 200 LSASLAAAEAFVLSGRNLNSRTYRQNSDNTSRTNWPSVSNFGTTNTNNSNTNS 259
DB 232 LSASLAAAEAFVLSGRNLNSRTYRQNSDNTSRTNWPSVSNFGTTNTNNSNTNS 291
OY 260 QASOSNDIDSLKQLYKPLSORHVESDGLVDPDQITSRITAKGVAVPHGDHYHIPYQ 319
DB 292 QASOSNDIDSLKQLYKPLSORHVESDGLVDPDQITSRITAKGVAVPHGDHYHIPYQ 351
OY 320 MSELERFARITPLPYRSNHWVPSRPPSPQPTPEPSGQPPAPNLIKIDNSLSVQL 379
DB 352 MSELERFARITPLPYRSNHWVPSRPPSPQPTPEPSGQPPAPNLIKIDNSLSVQL 411
OY 380 VRKVEGVVFEKGIISRYVFAKDLPSFTYKNLESKLKSQESVSHLTAKKEVAPRDOEF 439
DB 412 VRKVEGVVFEKGIISRYVFAKDLPSFTYKNLESKLKSQESVSHLTAKKEVAPRDOEF 471
OY 440 YDKAVNLTLEAHKALFNKNGRNSDFOALDKLIERLNDSTNKEKLVDDLAFAPITHE 499
DB 472 YDKAVNLTLEAHKALFNKNGRNSDFOALDKLIERLNDSTNKEKLVDDLAFAPITHE 531

```

```

OY 500 RLGRPNQISEYTEDEBVAIAQLADKYTTSDGYIPDEHDIISDEGDYVTPPHNGSHMIGKD 559
DB 532 RLGRPNQISEYTEDEBVAIAQLADKYTTSDGYIPDEHDIISDEGDYVTPPHNGSHMIGKD 591
OY 560 SLSDKEKVAQAAYTEKGIILPPSPDADVKANPTGDSAAAIYNRVKGKRIPLVRLPYVE 619
DB 592 SLSDKEKVAQAAYTEKGIILPPSPDADVKANPTGDSAAAIYNRVKGKRIPLVRLPYVE 651
OY 620 HTVEVKGNLIIIPKDHVHNIKFAFPDHTYKANGYTLDELPTATIKYVEHPERPHSN 679
DB 652 HTVEVKGNLIIIPKDHVHNIKFAFPDHTYKANGYTLDELPTATIKYVEHPERPHSN 711
OY 680 DGMGNASEHVLGKGDHSEDPKNFKADEEPVEETPAEBEVQVETKEVAQLKEAEVLLA 739
DB 712 DGMGNASEHVLGKGDHSEDPKNFKADEEPVEETPAEBEVQVETKEVAQLKEAEVLLA 771
OY 740 KVTDSLLKANATETTLAGLRNNLTLOIMDNNSIMAEAEKLLALKGSNPSSVSKEKIN 796
DB 772 KVTDSLLKANATETTLAGLRNNLTLOIMDNNSIMAEAEKLLALKGSNPSSVSKEKIN 828

RESULT 4
Q9AG74 PRELIMINARY; PRT; 844 AA.
AC Q9AG74:
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Phpa.
GN PHPA.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2124685; PubMed=11349048;
RA Zhang Y., Masi A.W., Barniak V., Mountzouros K., Hostetter M.K.,
RA Green B.A.;
RT "Recombinant Phpa Protein, a Unique Histidine Motif-Containing Protein
RT from Streptococcus pneumoniae, Protects Mice against Intranasal
RT Pneumococcal Challenge.";
RL Infect. Immun. 69:3827-3836(2001).
DR EMBL: AF340221; AAK26629.1; -.
DR InterPro: IPR006270; Strep_his_triad.
DR Triflame; TIGR01363; Strep_his_triad; 2
SQ SEQUENCE 844 AA; 94769 MW; D738A55290FF8902 CRC64;

Query Match 13.4%; Score 107; DB 2; Length 844;
Best Local Similarity 100.0%; Pred. No. 1,2e-96;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 102 KGGYIVKDGKYVYLKDAHADNVTKKEINRQOEHSOHRGCTPRNDGAVALARSG 161
DB 122 KGGYIVKDGKYVYLKDAHADNVTKKEINRQOEHSOHRGCTPRNDGAVALARSG 181
OY 162 RYTTDDGYIFNADIIEDTGDAYIVPHGDHYHIPKNELSASELAAA 208
DB 182 RYTTDDGYIFNADIIEDTGDAYIVPHGDHYHIPKNELSASELAAA 228

RESULT 5
Q8CWR4 PRELIMINARY; PRT; 855 AA.
AC Q8CWR4:
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Histidine motif-containing protein.
GN PHPA OR SPRI060.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.

```

OK NCBI_TaxId=171101;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21429245; PubMed=11544234;
 RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., BURGESS S.,
 RA DeHoff B.S., Estrem S.T., Filtz L., Fu D.-J., Fuller W., Geringer C.,
 RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
 RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
 RA McArthur S.M., McHenry M., Mcleaster K., Mundy C.W., Nicas T.I.,
 RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rokey P.,
 RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
 RA Zook C.A., Baltz R.H., Jaskunas S.R., Kosteck P.R. Jr., Skatrud P.L.,
 RA Glass J.I.,
 RT "Genome of the bacterium Streptococcus pneumoniae strain R6."
 RL J. Bacteriol. 183:5709-5717(2001).
 DR EMBL; AE008479; AAK99864.1; -
 KW Complete proteome.
 SQ SEQUENCE 855 AA; 96177 MW; 4350E82A3F97089A CRC64;

Query Match 13.4%; Score 107; DB 16; Length 855;
 Best Local Similarity 100.0%; Pred. No. 1.2e-96;
 Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 KGGYIKVKGKYYVYIKDAADNVTKEINRQKQHSQHRBGTPRNDGAVALARSG 161
 DB 133 KGGYIKVKGKYYVYIKDAADNVTKEINRQKQHSQHRBGTPRNDGAVALARSG 192

QY 162 RYTTDDGYFENASDIEDTDGAYIVPHGDHYHYIPKNELSASELAA 208
 DB 193 RYTTDDGYFENASDIEDTDGAYIVPHGDHYHYIPKNELSASELAA 239

RESULT 6

Q9ANY3 PRELIMINARY; PRT; 819 AA.

AC Q9ANY3;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Pneumococcal histidine triad protein B precursor (Fragment).
 GN PHTP.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxId=1313;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=21101045; PubMed=11159990;
 RA Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,
 RA Dornitzer M., Dagan R., Brewah Y.A., Barren P., Lathigra R.,
 RA Langemann S., Koenig S., Johnson S.;
 RT "Identification and characterization of a novel family of pneumococcal
 RT proteins (the Pht family) that are protective against sepsis."
 RL Infect. Immun. 69:949-958(2001).
 DR EMBL; AF318954; AAK06759.1; -
 DR InterPro; IPR006270; strep_his_triad.
 DR TIGRFAMS; TIGR01363; strep_his_triad; 2.
 KW Signal.
 FT SIGNAL. 1 29 POTENTIAL.
 FT NON_TER 819 819
 SQ SEQUENCE 819 AA; 92108 MW; E602CFC16CC28A5F CRC64;

Query Match 7.5%; Score 60; DB 2; Length 819;
 Best Local Similarity 100.0%; Pred. No. 4.6e-50;
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 ENLTPDEVSKREGINAEOIVIKITDGYVTSQGHYHYNGKVPYDAIISEELMKDPY 90
 DB 52 ENLTPDEVSKREGINAEOIVIKITDGYVTSQGHYHYNGKVPYDAIISEELMKDPY 111

RESULT 7

Q970M9

ID Q970M9 PRELIMINARY; PRT; 819 AA.

AC Q970M9;
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Conserved domain protein.
 GN SP1174.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxId=1313;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=TIGR4;
 RX MEDLINE=21357209; PubMed=11463916;
 RA Tetelzin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
 RA Peterson S., Heidelberg J., Debay R.T., Hatt D.H., Dodson R.J.,
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
 RA Holtzapfel E., Khouri H., Wolf A.M., Utecherback T.R., Hansen C.L.,
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
 RT "Complete genome sequence of a virulent isolate of Streptococcus
 RT pneumoniae."
 RL EMBL; AE007418; AAK75283.1; -

DR TIGR; SP1174; -
 DR InterPro; IPR006270; strep_his_triad.
 DR TIGRFAMS; TIGR01363; strep_his_triad; 2.
 KW Complete proteome.
 SQ SEQUENCE 819 AA; 92228 MW; 43852B72E8163BDE CRC64;

Query Match 7.5%; Score 60; DB 16; Length 819;
 Best Local Similarity 100.0%; Pred. No. 4.6e-50;
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 ENLTPDEVSKREGINAEOIVIKITDGYVTSQGHYHYNGKVPYDAIISEELMKDPY 90
 DB 52 ENLTPDEVSKREGINAEOIVIKITDGYVTSQGHYHYNGKVPYDAIISEELMKDPY 111

QY 52 ENLTPDEVSKREGINAEOIVIKITDGYVTSQGHYHYNGKVPYDAIISEELMKDPY 111

QY 52 ENLTPDEVSKREGINAEOIVIKITDGYVTSQGHYHYNGKVPYDAIISEELMKDPY 111

QY 52 ENLTPDEVSKREGINAEOIVIKITDGYVTSQGHYHYNGKVPYDAIISEELMKDPY 111

QY 52 ENLTPDEVSKREGINAEOIVIKITDGYVTSQGHYHYNGKVPYDAIISEELMKDPY 111

QY 52 ENLTPDEVSKREGINAEOIVIKITDGYVTSQGHYHYNGKVPYDAIISEELMKDPY 111

QY 52 ENLTPDEVSKREGINAEOIVIKITDGYVTSQGHYHYNGKVPYDAIISEELMKDPY 111

QY 52 ENLTPDEVSKREGINAEOIVIKITDGYVTSQGHYHYNGKVPYDAIISEELMKDPY 111

QY 52 ENLTPDEVSKREGINAEOIVIKITDGYVTSQGHYHYNGKVPYDAIISEELMKDPY 111

QY 52 ENLTPDEVSKREGINAEOIVIKITDGYVTSQGHYHYNGKVPYDAIISEELMKDPY 111

QY 52 ENLTPDEVSKREGINAEOIVIKITDGYVTSQGHYHYNGKVPYDAIISEELMKDPY 111

QY 52 ENLTPDEVSKREGINAEOIVIKITDGYVTSQGHYHYNGKVPYDAIISEELMKDPY 111

QY 52 ENLTPDEVSKREGINAEOIVIKITDGYVTSQGHYHYNGKVPYDAIISEELMKDPY 111

QY 52 ENLTPDEVSKREGINAEOIVIKITDGYVTSQGHYHYNGKVPYDAIISEELMKDPY 111

```

RA Holtzapfle E., Khouri H., Wolf A.M., Uterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiolini S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RL Science 293:498-506(2001).
DR EMBL: AF318955; AAK06760.1; -.
DR EMBL: AE007403; AAK75120.1; -.
DR TIGR: SP1003; -.
DR InterPro: IPR006270; Strep_his_triad.
DR SignalFAMS: TIGR01363; strep_his_triad; 2.
KW Signal, Hypothetical protein; Complete proteome.
FT SIGNAL 1 29
FT NON TER 839 839
SQ SEQUENCE 839 AA; 93672 MW; 713B180D5E03BDCA CRC64;

Query Match 7.5%; Score 60; DB 16; Length 839;
Best Local Similarity 100.0%; Pred. No. 4.7e-50;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 31 ENLTPDEVSKRGINAEOIVIKITDGYVTSFGDHYHYNGKVPYDAIISELLMKDPNY 90
Db 52 ENLTPDEVSKRGINAEOIVIKITDGYVTSFGDHYHYNGKVPYDAIISELLMKDPNY 111

RESULT 9
Q8D008 PRELIMINARY; PRT; 853 AA.
AC Q8D008;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Pneumococcal histidine triad protein D.
GN PHTD OR SPR0907.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA Dehoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA Leblanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushita P.,
RA McAhren S.M., McHenry M., Mcleaster K., Mundy C.W., Nicas T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostock P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6."
RL J. Bacteriol. 183:5709-5717(2001).
DR EMBL: AE008464; AAK99711.1; -.
KW Complete proteome.
SQ SEQUENCE 853 AA; 95225 MW; 98F06A1EDE90CFS CRC64;

Query Match 7.5%; Score 60; DB 16; Length 853;
Best Local Similarity 100.0%; Pred. No. 4.7e-50;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 31 ENLTPDEVSKRGINAEOIVIKITDGYVTSFGDHYHYNGKVPYDAIISELLMKDPNY 90
Db 52 ENLTPDEVSKRGINAEOIVIKITDGYVTSFGDHYHYNGKVPYDAIISELLMKDPNY 111

RESULT 10
Q9ANY1 PRELIMINARY; PRT; 1039 AA.
AC Q9ANY1;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)

```

```

DE Pneumococcal histidine triad protein E precursor (Hypothetical protein
DE SP1004).
GN PHTD OR SP1004.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21101045; PubMed=11159990;
RA Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,
RA Domitzier M., Dagan R., Brewah Y.A., Barren P., Lathigra R.,
RA Langemann S., Koenig S., Johnson S.;
RT "Identification and characterization of a novel family of pneumococcal
RT proteins (the Pht family) that are protective against sepsis."
RL Infect. Immun. 69:949-958(2001).
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN=TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tetteijn H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Rådune D.,
RA Holtzapfle E., Khouri H., Wolf A.M., Uterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiolini S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RL Science 293:498-506(2001).
DR EMBL: AF318955; AAK06761.1; -.
DR EMBL: AE007403; AAK75121.1; -.
DR TIGR: SP1004; -.
DR InterPro: IPR006270; Strep_his_triad.
DR SignalFAMS: TIGR01363; strep_his_triad; 4.
KW Signal, Hypothetical protein; Complete proteome.
FT SIGNAL 1 29
FT NON TER 114631 114631
SQ SEQUENCE 1039 AA; 114631 MW; 81A563C080625C4 CRC64;

Query Match 4.0%; Score 32; DB 16; Length 1039;
Best Local Similarity 100.0%; Pred. No. 3.2e-22;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 46 AEOIVIKITDGYVTSFGDHYHYNGKVPYDA 77
Db 67 AEOIVIKITDGYVTSFGDHYHYNGKVPYDA 98

RESULT 11
Q8D007 PRELIMINARY; PRT; 1039 AA.
AC Q8D007;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Pneumococcal histidine triad protein E.
GN PHTD OR SPR0908.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA Dehoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA Leblanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushita P.,
RA McAhren S.M., McHenry M., Mcleaster K., Mundy C.W., Nicas T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostock P.R. Jr., Skatrud P.L.,

```

RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6."
RL J. Bacteriol. 183:5709-5717(2001).
DR EMBL; AE008464; AAK9712.1; -
KW Complete proteome.
SQ SEQUENCE 1039 AA, 114625 MW, 05CC26D2028F551 CRC64;

Query Match
Best Local Similarity 100.0%; Score 32; DB 16; Length 1039;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 AEOIVIKITDQGVYTSHGHHYNGKVPYDA 77
DB 67 AEOIVIKITDQGVYTSHGHHYNGKVPYDA 98

RESULT 12

Q8E338 PRELIMINARY; PRT; 481 AA.
AC Q8E338;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN GBS1925.
OS Streptococcus agalactiae (serotype III).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_Taxid=216495;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEM316 / Serotype III;
RX MEDLINE=2242508; PubMed=12354221;
RA Glaeser P., Ruzniok C., Buchrieser C., Chevalier F., Frangoul L.,
MSadek T., Zouine M., Couve E., Lailoui L., Poyart C., Tlieu-Cuot P.,
Kunst F.;
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
RT invasive neonatal disease."
RL Mol. Microbiol. 45:1499-1513(2002).
DR EMBL; AL766854; CAD47584.1; -
KW Sagalier; gbs1925; -
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 481 AA; 53326 MW, DFD0453DA923BE CRC64;

Query Match
Best Local Similarity 100.0%; Score 27; DB 16; Length 481;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 GYVTSHGHHYNGKVPYDAISEEL 83
DB 81 GYVTSHGHHYNGKVPYDAISEEL 107

RESULT 13

Q9AE21 PRELIMINARY; PRT; 289 AA.
AC Q9AE21;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 32.0 kDa protein (Fragment).
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_Taxid=1311;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5531;
RX MEDLINE=21172873; PubMed=11274116;
RA Granlund M., Michel F., Norgren M.;
RT "Mutually exclusive distribution of IS1548 and GBS1, an active group
RT II intron identified in human isolates of group b streptococci."
RT J. Bacteriol. 183:2560-2569(2001).

DR EMBL; AJ290952; CAC35985.1; -
DR InterPro; IPR006270; strep_his_triad.
DR TIGRPFMS; TIGR01363; strep_his_triad; 1.
KW Hypothetical protein.
FT NON_TER 289
SQ SEQUENCE 289 AA; 32043 MW, A15A8588EA8140E4 CRC64;

Query Match
Best Local Similarity 100.0%; Score 17; DB 2; Length 289;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YNGKVPYDAISEELM 85
DB 92 YNGKVPYDAISEELM 108

RESULT 14

Q9ZHG7 PRELIMINARY; PRT; 822 AA.
AC Q9ZHG7;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 92.4 kDa protein.
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_Taxid=1311;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R268;
RX MEDLINE=99115568; PubMed=9916102;
RA Spellerberg B., Rozdzinski E., Martin S., Weber-Heymann J.,
RA Schmitz N., Luetticken R., Podbielski A.,
RT "Lmo, a protein with similarities to the Lrai adhesin family, mediates
RT attachment of Streptococcus agalactiae to human laminin."
RL Infect. Immun. 67:871-878(1999).
DR EMBL; AF062533; AAD13797.1; -
DR InterPro; IPR006270; strep_his_triad.
DR TIGRPFMS; TIGR01363; strep_his_triad; 4.
KW Hypothetical protein.
SQ SEQUENCE 822 AA; 92386 MW, 80E4EDF313481F98 CRC64;

Query Match
Best Local Similarity 100.0%; Score 17; DB 2; Length 822;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YNGKVPYDAISEELM 85
DB 92 YNGKVPYDAISEELM 108

RESULT 15

Q8E4U1 PRELIMINARY; PRT; 822 AA.
AC Q8E4U1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN GBS1306.
OS Streptococcus agalactiae (serotype III).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_Taxid=216495;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEM316 / Serotype III;
RX MEDLINE=2242508; PubMed=12354221;
RA Glaeser P., Ruzniok C., Buchrieser C., Chevalier F., Frangoul L.,
MSadek T., Zouine M., Couve E., Lailoui L., Poyart C., Tlieu-Cuot P.,
Kunst F.;
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing

RT invasive neonatal disease.";
 RL Microbiol. 45:1499-1513 (2002).
 DR EMBL, AL766850; CAD46965.1; -.
 DR Sagelact; gbs1306; -.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 822 AA; 92393 MW; 2929A97C8AFCDF78F CRC64;

Query Match 2.1%; Score 17; DB 16; Length 822;
 Best Local Similarity 100.0%; Pred. No. 1.9e-07;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YNGKVPYDAIISEBLM 85
 |||||
 DB 92 YNGKVPYDAIISEBLM 108

RESULT 16

Q8D281 PRELIMINARY; PRT; 822 AA.

AC Q8D281;
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DE Streptococcal histidine triad family protein.
 GN SAG1233.
 OS Streptococcus agalactiae (serotype V).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=216466;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=2603 V/R / Serotype V;
 RX MEDLINE=22222988; PubMed=12200547;

RA Tetsellin H., Madsen I.T., Nelson K.E., Margalit I., Read T.D.,
 RA Wesels M.R., Paulsen I.T., Beaman M.J., Brinkac L.M., Daugherty S.C.,
 RA Madoff L.C., Wolf A.M., Koianay J.F., Madupu R., Lewis M.R.,
 RA DeBoy R.T., Durkin A.S., Scanlan D., Khouri H., Mulligan S.,
 RA Radu D., Fedorova N.B., Scanlan D., Khouri H., Mulligan S.,
 RA Caray H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
 RA Jacobini E.T., Breton C., Galli G., Mariani M., Vanni F., Malone D.,
 RA Rinaudo D., Rappaport R., Telford J.L., Kasper D.L., Grandi G.,
 RA Fraser C.M.;
 RA "Complete genome sequence and comparative genomic analysis of an
 RT emerging human pathogen, serotype V Streptococcus agalactiae.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396 (2002).
 DR EMBL; AE014246; AAN011.1; -.
 DR TIGR; SAG1233; -.
 KW Complete proteome.
 SQ SEQUENCE 822 AA; 92400 MW; 8CC8DF316727F98 CRC64;

Query Match 2.1%; Score 17; DB 16; Length 822;
 Best Local Similarity 100.0%; Pred. No. 1.9e-07;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YNGKVPYDAIISEBLM 85
 |||||
 DB 92 YNGKVPYDAIISEBLM 108

RESULT 17

Q8N282 PRELIMINARY; PRT; 823 AA.

AC Q8N282;
 DT 01-OCT-2002 (TREMBLrel. 22, Created)
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
 DE Hypothetical protein spy18_2072.
 GN SPY18_2072.
 OS Streptococcus pyogenes (serotype M18).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=186103;
 RN (1)

RP SEQUENCE FROM N.A.
 RC STRAIN=MGAS8232 / Serotype M18;
 RX MEDLINE=21927593; PubMed=11917108;

RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
 RA Sylva G.L., Sturtevant D.E., Ricklets S.M., Porcella S.F.,
 RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
 RA Kapur V., Daly J.A., Vasey L.G., Musser J.M.;
 RT "Genome sequence and comparative microarray analysis of serotype M18
 RT group A Streptococcus strains associated with acute rheumatic fever
 RT outbreaks.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673 (2002).
 DR EMBL; AE010110; AAL98543.1; -.
 DR InterPro; IPR006270; Strep_his_triad.
 DR TIGRFAMs; TIGR01363; strep_his_triad; 4.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 823 AA; 92585 MW; C79E1EB30CEDDF0C CRC64;

Query Match 2.1%; Score 17; DB 16; Length 823;
 Best Local Similarity 100.0%; Pred. No. 1.9e-07;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YNGKVPYDAIISEBLM 85
 |||||
 DB 92 YNGKVPYDAIISEBLM 108

RESULT 18

Q8K5Q1 PRELIMINARY; PRT; 823 AA.

AC Q8K5Q1;
 DT 01-OCT-2002 (TREMBLrel. 22, Created)
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Histidine triad protein.
 GN SPY18_1724.
 OS Streptococcus pyogenes (serotype M3).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=186466;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=MGAS315 / Serotype M3;
 RX MEDLINE=22133808; PubMed=12122206;

RA Beres S.B., Sylva G.L., Barbican K.D., lei B., Hoff J.S.,
 RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,
 RA Campbell D.S., Smith T.M., McCormack J.K., Leung D.Y.M.,
 RA Schlievert P.M., Musser J.M.;
 RT "Genome sequence of a serotype M3 strain of group A Streptococcus:
 RT phage-encoded toxins, the high-virulence phenotype, and clone
 RT emergence.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083 (2002).
 DR EMBL; AE014169; AAM80331.1; -.
 DR InterPro; IPR006270; Strep_his_triad.
 DR TIGRFAMs; TIGR01363; strep_his_triad; 1.
 KW Complete proteome.
 SQ SEQUENCE 823 AA; 92431 MW; 8C6CBD517A2DD616 CRC64;

Query Match 2.1%; Score 17; DB 16; Length 823;
 Best Local Similarity 100.0%; Pred. No. 1.9e-07;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YNGKVPYDAIISEBLM 85
 |||||
 DB 92 YNGKVPYDAIISEBLM 108

RESULT 19

Q93GTS PRELIMINARY; PRT; 825 AA.

AC Q93GTS;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

```

DE Histidine triad protein of group A streptococci.
GN HTPA.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxId=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SSI-9;
RA Teroo Y., Kawabata S., Hamada S.;
RT "Characterization of a novel histidine triad protein of group A
RT streptococci."
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB073859; BAB71774.1; -.
DR InterPro; IPR006270; Strep_his_triad.
DR TIGRFAMs; TIGR01363; strep_his_triad; 4.
SQ SEQUENCE 825 AA; 92623 MW; DEAECC199181DFB CRC64;

Query Match
Best Local Similarity 2.1%; Score 17; DB 2; Length 825;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YNGKVPYDAIISEELM 85
DB 92 YNGKVPYDAIISEELM 108

RESULT 20
Q99XV4 PRELIMINARY; PRT; 825 AA.
AC Q99XV4;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein SPY2006.
GN SPY2006.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxId=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferrerli J.J., McShan W.M., Ajdic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL; AE006623; AAK34688.1; -.
DR InterPro; IPR006270; Strep_his_triad.
DR TIGRFAMs; TIGR01363; strep_his_triad; 4.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 825 AA; 92649 MW; 57DF509656D50F4 CRC64;

Query Match
Best Local Similarity 2.1%; Score 17; DB 16; Length 825;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YNGKVPYDAIISEELM 85
DB 92 YNGKVPYDAIISEELM 108

RESULT 21
Q8DQ06 PRELIMINARY; PRT; 182 AA.
AC Q8DQ06;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Pneumococcal histidine triad protein E, truncation.

```

```

GN PHE-TRUNCATION OR SPRO10.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxId=171101;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszczyk L.C., Burgett S.,
RA Dehoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glase J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Letkowitz E.J., Lu J., Matsushima P.,
RA McAnren S.M., McHenry M., Mcleaster K., Mundy C.W., Nicas T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rokey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
RA Glase J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
DR EMBL; AE008464; AAK99714.1; -.
KW Complete proteome.
SQ SEQUENCE 182 AA; 20145 MW; 29E12D653313A343 CRC64;

Query Match
Best Local Similarity 1.3%; Score 10; DB 16; Length 182;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 188 HGDHYHYIPK 197
DB 35 HGDHYHYIPK 44

RESULT 22
Q9CL01 PRELIMINARY; PRT; 791 AA.
AC Q9CL01;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein PM1448.
GN PM1448.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Pasteurella.
OX NCBI_TaxId=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pm70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whitlam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL; AE006181; AAK03532.1; -.
DR HSP; P05055; ISRO.
DR InterPro; IPR003583; HHH_1.
DR InterPro; IPR003029; SL_1.
DR InterPro; IPR006641; YGFC.
DR Pfam; PF0575; SL_1.
DR SMART; SM00278; Hhh1; 2.
DR SMART; SM00316; SL_1.
DR SMART; SM00732; YGFC; 1.
DR PROSITE; PS50126; SL_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 791 AA; 87498 MW; 8C5D1C5401255C0 CRC64;

Query Match
Best Local Similarity 1.1%; Score 9; DB 16; Length 791;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 201 SASSELAANE 209
DB 443 SASSELAANE 451

```

RESULT 23

Q8B9W0 PRELIMINARY; PRT; 71 AA.
 ID Q8B9W0
 AC Q8B9W0
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein.
 OS Rachiplusia on multiple nucleopolyhedrovirus.
 CC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
 CC Nucleopolyhedrovirus.
 NC NCB1_TaxID=803166;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bonning B.C., Harrison R.L.;
 RT "The Rachiplusia on multiple nucleopolyhedrovirus genome sequence."
 RL Submitted (Aug-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AY145471; AAN28126.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 71 AA; 8522 MW; 924875A256D163BD CRC64;

Query Match 1.0%; Score 8; DB 12; Length 71;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 EBNRKO 137
 DB 20 EBNRKO 27

RESULT 24

Q8EHD8 PRELIMINARY; PRT; 147 AA.
 ID Q8EHD8
 AC Q8EHD8
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Conserved hypothetical protein.
 GN SO1287.
 OS Shewanella oneidensis.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
 CC Alteromonadaceae; Shewanella.
 OC NCB1_TaxID=70863;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=MR-1;
 RX MEDLINE=22297686; PubMed=12368813;
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
 RA Read T.D., Eisen J.A., Seshadri R., Ward N., Meche B., Clayton R.A.,
 RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
 RA Deboy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
 RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
 RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
 RA Mueller J., Khouri H., Gill J., Uterback T.R., McDonald L.A.,
 RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.;
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
 Shewanella oneidensis."
 RL Nat. Biotechnol. 20:1118-1123(2002).
 DR EMBL; AE015573; AAN54354.1; -
 DR TIGR; SO1287; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 147 AA; 15926 MW; 659276DA5626367 CRC64;

Query Match 1.0%; Score 8; DB 16; Length 147;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 SELAAEA 210
 DB 73 SELAAEA 80

RESULT 25

O811B9 PRELIMINARY; PRT; 152 AA.
 ID O811B9
 AC O811B9
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein.
 GN Pfl1.0255.
 OS Plasmodium falciparum (isolate 3D7).
 CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 CC NCB1_TaxID=36329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3D7;
 RX MEDLINE=22255705; PubMed=12368864;
 RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
 RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
 RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
 RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Anguillo S.,
 RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
 RA Martin D.M.A., Fairhead A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
 RA McRadden G.I., Cummings L.M., Sudramanian G.M., Mungall C.,
 RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
 RA Fraser C.M., Barrall B.;
 RT "Genome sequence of the human malaria parasite Plasmodium
 falciparum."
 RL Nature 419:498-511(2002).
 DR EMBL; AE014839; AAN35839.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 152 AA; 17868 MW; CAEB34CE325F7800 CRC64;

Query Match 1.0%; Score 8; DB 5; Length 152;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 VKNGNLI 631
 DB 69 VKNGNLI 76

RESULT 26

P91742 PRELIMINARY; PRT; 154 AA.
 ID P91742
 AC P91742
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE Transposase (Fragment).
 OS Hydra attenuata (Hydra) (Hydra vulgaris).
 CC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Anthomedusae;
 CC Hydridae; Hydra.
 OC NCB1_TaxID=6087;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSOM=mariner;
 RX MEDLINE=97327060; PubMed=9183847;
 RA Robertson H.M.;
 RT "Multiple Mariner transposons in flatworms and hydras are related to
 those of insects."
 RL J. Hered. 88:195-201(1997).
 DR EMBL; U51185; AAB61389.1; -
 DR InterPro; IPR001888; Transposase_1.
 DR Pfam; PF01359; Transposase_1; 1.
 FT NON_TER 1
 FT NON_TER 154
 SQ SEQUENCE 154 AA; 18179 MW; 955FB7091FA34F57 CRC64;

Query Match 1.0%; Score 8; DB 5; Length 154;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 470 LLERLND 477

Db 103 LBERLNDE 110

RESULT 27

ID 044095 PRELIMINARY; PRT; 181 AA.
 AC 044095;
 DT 01-JUN-1998 (TEMBLrel. 06, Created)
 DT 01-JUN-1998 (TEMBLrel. 06, Last sequence update)
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
 DE Adenine phosphoribosyltransferase (Fragment).
 GN APT.
 OS Drosophila subobscura (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OK NCBI_TaxId=7241;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Zeng L.-W., Comeran J.M., Chen B., Kreitman M.,
 RL Genetics 0:0-0(1997).
 CC -1- SIMILARITY: BELONGS TO THE PURINE/PYRIMIDINE
 CC PHOSPHORIBOSYLTRANSFERASE FAMILY.
 DR EMBL; AF025800; AAB87885.1; -.
 DR FlyBase; FBgn0023247; Dsub\Aptc.
 DR InterPro; IPR005764; Ade.phapb.trans.
 DR InterPro; IPR002375; Pr/PY_rp.transf.
 DR InterPro; IPR000836; PRTtransferase.
 DR Pfam; PF00156; Pribosyltran; 1.
 DR TIGRFAMs; TIGR01090; apt; 1.
 DR PROSITE; PS00103; PUR_PYR_PR_TRANSFER; 1.
 DR GlycoSyLtransferase; Transferase.
 FT NON_TER 181
 SQ SEQUENCE 181 AA; 1986 MW; F34786596C31F726 CRC64;
 Query Match 1.0%; Score 8; DB 5; Length 181;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 377 SOLVRKVG 384
 Db 143 SOLVRKVG 150
 RESULT 28
 Q9NCF0 PRELIMINARY; PRT; 270 AA.
 ID Q9NCF0;
 AC Q9NCF0;
 DT 01-OCT-2000 (TEMBLrel. 15, Created)
 DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)
 DT 01-OCT-2002 (TEMBLrel. 22, Last annotation update)
 DE Beta-1,4-glucanase 2 (Fragment).
 GN EG2.
 OS Cryptocercus cleavelandi.
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Orthopteroidea; Dictyoptera; Blattaria; Blattellidae;
 OC Cryptocercidae; Cryptocercus.
 OK NCBI_TaxId=119679;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA TISSUE=Salivary gland;
 RX MEDLINE=20360158; PubMed=10898984;
 RA Lo N., Tokuda G., Watanabe H., Rose H., Slaytor M., Maekawa K.,
 RA Bandi C., Noda H.;
 RT "Evidence from multiple gene sequences indicated that termites evolved
 RT from wood-feeding cockroaches.";
 RL Curr. Biol. 10:801-804(2000).
 DR EMBL; AF220589; AAF63720.1; -.
 DR HSSP; P42621; 1TFA.
 DR InterPro; IPR001701; Glyco_hydro_9.
 DR Pfam; PF00759; Glyco_hydro_9; 1.
 FT NON_TER 1

FT NON_TER 270 270
 SQ SEQUENCE 270 AA; 29875 MW; B03BF2A7ABF0547C CRC64;
 Query Match 1.0%; Score 8; DB 5; Length 270;
 Best Local Similarity 100.0%; Pred. No. 58;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 735 EVLLAKVT 742
 Db 223 EVLLAKVT 230
 RESULT 29
 Q22866 PRELIMINARY; PRT; 284 AA.
 ID Q22866;
 AC Q22866;
 DT 01-NOV-1996 (TEMBLrel. 01, Created)
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
 DE CetM (Y105EBB.1a protein).
 DE Y105EBB.1 OR TMV-1 OR Y105EBB.1A.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OK NCBI_TaxId=6239;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA STRAIN=Bristol N2;
 RX MEDLINE=95395840; PubMed=7666414;
 RA Kagawa H., Sugimoto K., Matsumoto H., Inoue T., Imadzu H., Takuwa K.,
 RA Sakabe Y.;
 RT "Genome structure, mapping and expression of the tropomyosin gene tmy-
 RT 1 of Caenorhabditis elegans.";
 RL J. Mol. Biol. 251:603-613(1995).
 RN (2)
 RP SEQUENCE FROM N.A.
 RA Sulston J.E.;
 RL Submitted (Oct-2001) to the EMBL/Genbank/DBJ databases.
 RN (3)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology.";
 RL Science 282:2012-2018(1998).
 DR EMBL; D38540; BAA07543.1; -.
 DR EMBL; AL132877; CAC70114.1; -.
 DR HSSP; P42639; 1CIG.
 DR WormBep; Y105EBB.1a; CE28782.
 DR InterPro; IPR000533; Tropomyosin.
 DR Pfam; PF00261; Tropomyosin; 1.
 DR PRINTS; PR00194; TROPOMYOSIN.
 DR PROSITE; PS00326; TROPOMYOSIN; 1.
 SQ SEQUENCE 284 AA; 33003 MW; 02271C870E23D2AB CRC64;
 Query Match 1.0%; Score 8; DB 5; Length 284;
 Best Local Similarity 100.0%; Pred. No. 61;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 727 VEAQLKEA 734
 Db 144 VEAQLKEA 151
 RESULT 30
 Q22865 PRELIMINARY; PRT; 284 AA.
 ID Q22865;
 AC Q22865;
 DT 01-NOV-1996 (TEMBLrel. 01, Created)
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
 DE CetM1.
 GN TMV-1.

OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OK NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=95395840; PubMed=7666414;
 RA Kagawa H., Sugimoto K., Matsumoto H., Inoue T., Imadzu H., Takuwa K.,
 RA Sakube Y.;
 RT "Genome structure, mapping and expression of the tropomyosin gene tmy-
 1 of Caenorhabditis elegans.";
 RL J. Mol. Biol. 251:603-613(1995).
 DR EMBL; D38539; BAA07540.1; -;
 DR HSSP; P42639; ICIG.
 DR InterPro; IPR000533; Tropomyosin.
 DR Pfam; PF00261; Tropomyosin.1.
 DR PRINTS; PR00194; TROPOMYOSIN.
 DR PROSITE; PS00326; TROPOMYOSIN; 1.
 SQ SEQUENCE 284 AA; 32951 MW; 80271C870E26C5E CRC64;
 Query Match 1.0%; Score 8; DB 5; Length 284;
 Best Local Similarity 100.0%; Pred. No. 61;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 727 VEAQLKEA 734
 DB 144 VEAQLKEA 151
 RESULT 31
 Q27284 PRELIMINARY; PRT; 284 AA.
 AC Q27284;
 DT 01-NOV-1996 (TRENBLREL. 01, Created)
 DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
 DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)
 DE CEMIII (Y105EBB.1D protein).
 GN Y105EBB.1 OR TMY-1 OR Y105EBB.1D.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OK NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RX MEDLINE=95395840; PubMed=7666414;
 RA Kagawa H., Sugimoto K., Matsumoto H., Inoue T., Imadzu H., Takuwa K.,
 RA Sakube Y.;
 RT "Genome structure, mapping and expression of the tropomyosin gene tmy-
 1 of Caenorhabditis elegans.";
 RL J. Mol. Biol. 251:603-613(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Sulston J.E.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 investigating biology.";
 RL Science 282:2012-2018(1998).
 DR EMBL; D38541; BAA07544.1; -;
 DR EMBL; D38539; BAA07541.1; -;
 DR EMBL; AL132877; CAC70112.1; -;
 DR HSSP; P42639; ICIG.
 DR WormPep; Y105EBB.1D; CE29060.
 DR InterPro; IPR000533; Tropomyosin.
 DR Pfam; PF00261; Tropomyosin.1.
 DR PRINTS; PR00194; TROPOMYOSIN.
 DR PROSITE; PS00326; TROPOMYOSIN; 1.
 KW Alternative splicing.

SQ SEQUENCE 284 AA; 32936 MW; D8D32CC6FB4E78F7 CRC64;
 Query Match 1.0%; Score 8; DB 5; Length 284;
 Best Local Similarity 100.0%; Pred. No. 61;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 727 VEAQLKEA 734
 DB 144 VEAQLKEA 151
 RESULT 32
 Q46010 PRELIMINARY; PRT; 294 AA.
 AC Q46010;
 DT 01-JUN-1998 (TRENBLREL. 06, Created)
 DT 01-JUN-1998 (TRENBLREL. 06, Last sequence update)
 DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)
 DE ZK228.4.
 GN ZK228.4.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OK NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Baaham V.M.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 investigating biology.";
 RL Science 282:2012-2018(1998).
 DR EMBL; Z82086; CAB0496.1; -;
 DR WormPep; ZK228.4; CB16713.
 DR InterPro; IPR000182; GCN5acetyltransf.
 DR Pfam; PF00583; Acetyltransf.1.
 SQ SEQUENCE 294 AA; 34067 MW; D91875015FB09976 CRC64;
 Query Match 1.0%; Score 8; DB 5; Length 294;
 Best Local Similarity 100.0%; Pred. No. 63;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 11 TVKENRV 18
 DB 262 TVKENRV 269
 RESULT 33
 Q95031 PRELIMINARY; PRT; 301 AA.
 AC Q95031;
 DT 01-DEC-2001 (TRENBLREL. 19, Created)
 DT 01-DEC-2001 (TRENBLREL. 19, Last sequence update)
 DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)
 DE Y105EBB.1b protein.
 GN Y105EBB.1 OR Y105EBB.1B.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OK NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Sulston J.E.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 investigating biology.";

```
RL Science 282:2012-2018(1998).
DR EMBL; AL132877; CAC70113.1; -.
DR HSSP; P04268; 1IC2.
DR WormRep; Y105EBB.1b; CE29058.
DR InterPro; IPR000533; Tropomyosin.
DR Pfam; PF00261; Tropomyosin; 1.
DR PRINTS; PR00194; TROPOMYOSIN.
DR PROSITE; PS00326; TROPOMYOSIN; 1.
SQ SEQUENCE 301 AA; 34832 MW; B6E190C25171D32F CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 5; Length 301;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 727 VEAOLKEA 734
Db 161 VEAOLKEA 168

RESULT 34
Q8FLF7 PRELIMINARY; PRT; 319 AA.
AC Q8FLF7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Hypothetical protein.
GN CE3P025.
OS Corynebacterium efficiens.
OG Plasmid pCE3.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
ON NCBI_TaxID=152794;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RA Kawabayashi Y., Yamazaki J., Hino Y., Kikuchi H.;
RT "The entire sequence of plasmid maintained by Corynebacterium
RT efficiens YS-314."
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF005326; BAC19600.1; -.
KW Hypothetical protein; Plasmid; Complete proteome.
SQ SEQUENCE 319 AA; 32777 MW; 38DA547FCB648122 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 319;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 708 EPVEETPA 715
Db 267 EPVEETPA 274

RESULT 35
Q9P8N1 PRELIMINARY; PRT; 436 AA.
AC Q9P8N1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Cellobiohydrolase.
OS Coriolus versicolor.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Aphyllophorales; Coriolus.
ON NCBI_TaxID=57466;
RN [1]
RP SEQUENCE FROM N.A.
RA Novo C., Clemente A., Simoes F., Mendonca D., Matos J.;
RT "Coriolus versicolor cellobiohydrolase-like cDNA sequence."
DR EMBL; AF233583; AAF35251.1; -.
DR HSSP; P00725; 2CBH.
DR InterPro; IPR000254; CBD_fungal.
```

```
DR InterPro; IPR001524; Glyco_hydro_6.
DR Pfam; PF00734; CBM.1; 1.
DR Pfam; PF01341; Glyco_hydro_6; 1.
DR PRINTS; PR00733; GLYDRLASES.
DR ProDom; PD001821; CBD_fungal; 1.
DR ProDom; PD003733; Glyco_hydro_6; 1.
DR SMART; SM00236; fCBP. 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
DR PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
DR PROSITE; PS00656; GLYCOSYL_HYDROL_F6_2; 1.
KW Hydrolase.
SQ SEQUENCE 436 AA; 46170 MW; E0BC5AFBCA324D4 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 3; Length 436;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 741 VTDSLSKA 748
Db 107 VTDSLSKA 114

RESULT 36
Q9KS51 PRELIMINARY; PRT; 484 AA.
AC Q9KS51;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Multidrug resistance protein, putative.
GN VC1409.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
ON NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Ginn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bases S., Qin H., DiGregori I., Sellers P.,
RA McDonald L., Ueberlack T., Fleischmann R.D., Niemann W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
DR EMBL; AE004219; AAF94566.1; -.
DR TTGR; VC1409; -.
DR InterPro; IPR003423; OEP.
DR InterPro; IPR001440; TPR.
DR Pfam; PF02321; OEP; 2.
KW Complete proteome.
SQ SEQUENCE 484 AA; 52131 MW; 1C53BD09B969F203 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 484;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 SELAAEA 210
Db 169 SELAAEA 176

RESULT 37
Q80401 PRELIMINARY; PRT; 485 AA.
AC Q80401;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
```

DE SAR DNA binding protein (Fragment).
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OK NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Hatton D., Gray J.C.;
 RT "Rice cDNA encoding a protein homologous to a SAR DNA-binding protein
 from pea."
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB015431; BAA31260.1; -.
 DR Gramene; O80401; -.
 DR InterPro; IPR002687; Nop.
 DR Pfam; PF01798; Nop; 1.
 DR ProDom; PD004104; Nop; 1.
 DR NON_TER 485 485
 SQ SEQUENCE 485 AA; 5334 MW; 61A2986F7856D56D CRC64;

Query Match 1.0%; Score 8; DB 10; Length 485;
 Best Local Similarity 100.0%; Pred. No. 99;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 727 VEAQLKEA 734
 |||||
 Db 229 VEAQLKEA 236

RESULT 38

ID 0997D8 PRELIMINARY; PRT; 488 AA.
 AC 0997D8;
 DT 01-JUN-2001 (TRENBLREL. 17, Created)
 DT 01-JUN-2001 (TRENBLREL. 17, Last sequence update)
 DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
 DE PTS system, N-acetylglucosamine-specific IIABC component.
 GN PTA OR SA11726 OR SA1547.
 OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
 OS Staphylococcus aureus (strain N315).
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OK NCBI_TaxID=158878; 158879;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
 RX MEDLINE=21311952; PubMed=11418146;
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
 RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-Q., Ito T.,
 RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
 RA Mizutani-U Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
 RA Sekimizu K., Hiraoka H., Kuhara S., Goto S., Yabuzaki J.,
 RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
 RA Hatton D., Ogasawara N., Hayashi H., Hiramatsu K.;
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus
 aureus."
 RL Lancet 357:1225-1240(2001).
 DR EMBL; AP003363; BAB57888.1; -.
 DR EMBL; AP003134; BAB42815.1; -.
 DR HSSP; P05053; 11BA.
 DR InterPro; IPR001996; PTS_EIIB.
 DR InterPro; IPR003352; PTS_EIIC.
 DR Pfam; PF00367; PTS_EIIB; 1.
 DR Pfam; PF02378; PTS_EIIC; 1.
 DR ProDom; PD001476; PTS_EIIB; 1.
 DR PROSITE; PS01035; PTS_EIIB_CYS; 1.
 KW Complete proteome.
 SQ SEQUENCE 488 AA; 53086 MW; 4A669EF1ACB5ED4E CRC64;

Query Match 1.0%; Score 8; DB 16; Length 488;
 Best Local Similarity 100.0%; Pred. No. 99;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 150 NDGAVALA 157

Db 72 NDGAVALA 79
 |||||

RESULT 39

ID 08NM40 PRELIMINARY; PRT; 488 AA.
 AC 08NM40;
 DT 01-OCT-2002 (TRENBLREL. 22, Created)
 DT 01-OCT-2002 (TRENBLREL. 22, Last sequence update)
 DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)
 DE PTS system N-acetylglucosamine-specific IIABC component.
 GN PTA OR MW1668.
 OS Staphylococcus aureus (strain MW2).
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OK NCBI_TaxID=196620;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22040717; PubMed=12044378;
 RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
 RA Nagai Y., Iwama N., Asano K., Naito T., Kuroda H., Cui L.,
 RA Yamamoto K., Hiramatsu K.;
 RT "Genome and virulence determinants of high virulence community-
 acquired MRSA."
 RL Lancet 359:1819-1827(2002).
 DR EMBL; AP004827; BAB95533.1; -.
 DR InterPro; IPR001996; PTS_EIIB.
 DR InterPro; IPR003352; PTS_EIIC.
 DR Pfam; PF00367; PTS_EIIB; 1.
 DR Pfam; PF02378; PTS_EIIC; 1.
 DR ProDom; PD001476; PTS_EIIB; 1.
 DR PROSITE; PS01035; PTS_EIIB_CYS; 1.
 KW Complete proteome.
 SQ SEQUENCE 488 AA; 53098 MW; 52887CE71A4F39F CRC64;

Query Match 1.0%; Score 8; DB 16; Length 488;
 Best Local Similarity 100.0%; Pred. No. 99;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 150 NDGAVALA 157
 |||||
 Db 72 NDGAVALA 79

RESULT 40

ID 08CNW2 PRELIMINARY; PRT; 490 AA.
 AC 08CNW2;
 DT 01-MAR-2003 (TRENBLREL. 23, Created)
 DT 01-MAR-2003 (TRENBLREL. 23, Last sequence update)
 DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)
 DE PTS system N-acetylglucosamine-specific IIIC component.
 GN SE1403.
 OS Staphylococcus epidermidis.
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OK NCBI_TaxID=1282;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 12228;
 RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
 RA Chen Z., Wen Y.;
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB016748; AA005002.1; -.
 KW Complete proteome.
 SQ SEQUENCE 490 AA; 53646 MW; D1DE11C5EC289B3D CRC64;

Query Match 1.0%; Score 8; DB 16; Length 490;
 Best Local Similarity 100.0%; Pred. No. 99;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 150 NDGAVALA 157
 |||||
 Db 72 NDGAVALA 79

```

RESULT 41
ID 095LH9 PRELIMINARY; PRT; 673 AA.
AC 095LH9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Rab11 family interaction protein 1.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP MEDLINE=21486449; PubMed=11495908;
RX Hales C.M., Griner R., Hobby-Henderson K.C., Dorn M.C., Hardy D.,
RA Kumar R., Navarre R., Chan E.K., Lapierre L.A., Goldenring J.R.;
RT "Identification and characterization of a family of Rab11-interacting
RT proteins."
RL J. Biol. Chem. 276:39067-39075(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Griner R.D., Kumar R., Goldenring J.R.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF237668; AAL02425.2; -.
SQ SEQUENCE 673 AA; 69892 MW; A1B8FCDD6998B192 CRC64;

Query Match
Best Local Similarity 1.0%; Score 8; DB 6; Length 673;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 591 PTGDSAAA 598
DB 69 PTGDSAAA 76

RESULT 42
ID 09RSJ4 PRELIMINARY; PRT; 840 AA.
AC 09RSJ4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Protease, putative.
GN DR2130.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=R1;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Yang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.U., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarov K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1."
RL Science 286:1571-1577(1999).
DR EMBL; AE002047; AAF1680.1; -.
DR TIGR; DR2130; -.
DR InterPro; IPR001539; Peptidase_U32.
DR Pfam; PF01136; Peptidase_U32.1.
DR ProDom; PD004398; Peptidase_U32; 1.
KW Protease; Complete proteome.
SQ SEQUENCE 840 AA; 91640 MW; 3A978AD3A0742B88 CRC64;

```

```

Query Match
Best Local Similarity 1.0%; Score 8; DB 16; Length 840;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 DGLVDPDA 294
DB 332 DGLVDPDA 339

RESULT 43
ID 026077 PRELIMINARY; PRT; 945 AA.
AC 026077;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Helicase.
GN HP1553.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kervavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klein H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Richardson J., Zhou L., Khakhria E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wathey L., Mallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
DR EMBL; AE000653; AAD08593.1; -.
DR TIGR; HP1553; -.
DR InterPro; IPR000212; UvrD-helicase.
DR Pfam; PF00580; UvrD-helicase; 1.
KW Hypothetical protein; Helicase; Complete proteome.
SQ SEQUENCE 945 AA; 108880 MW; 6EE53868568EB22 CRC64;

Query Match
Best Local Similarity 1.0%; Score 8; DB 16; Length 945;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 499 ERLGKPN 506
DB 657 ERLGKPN 664

RESULT 44
ID 08TMX4 PRELIMINARY; PRT; 1233 AA.
AC 08TMX4;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE WD40-repeat containing protein.
GN MA2525.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., MacDonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Altnor D., Brown A.,

```

RA Allen N., Naylor J., Strange-Thomann N., Dearellano K., Johnson R.,
 RA Linton L., McEwan P., McKernan K., Talamas J., Titrrell A., Ye W.,
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
 RA Hedderich R., Ingram-Smith C., Kuettnner H.C., Krzycki J.A.,
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
 RA Perry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 RA Metcalf M.W., Birren B.;
 RT "The genome of *Methanoscoccus acetivorans* reveals extensive metabolic
 RT and physiological diversity.";
 RT Genome Res. 12:532-542(2002).
 DR EMBL; AE010946; AAM05908.1; -.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 16.
 DR SMART; SM00320; WD40; 16.
 DR PROSITE; PS00678; WD_REPEATS_1; 13.
 DR PROSITE; PS50082; WD_REPEATS_2; 16.
 DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
 KW Complete proteome.
 SQ SEQUENCE 1233 AA; 138558 MW; 49D207B08C3BB8A6 CRC64;

Query Match 1.0%; Score 8; DB 17; Length 1233;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 482 EKLVDL 489
 Db 236 EKLVDL 243

RESULT 45
 ID Q93HW0 PRELIMINARY; PRT; 1646 AA.
 AC Q93HW0;
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Endo-beta-N-acetylglucosaminidase D.
 GN ENDO D.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=2128261; PubMed=11388907;
 RA Muramatsu H., Tachikui H., Ushida H., Song X., Qiu Y., Yamamoto S.,
 RA Muramatsu T.;
 RT "Molecular Cloning and Expression of Endo-beta-N-acetylglucosaminidase
 RT D, Which Acts on the Core Structure of Complex Type Asparagine-Linked
 RT Oligosaccharides.";
 RT J. Biochem. 129:923-928(2001).
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOGLYCAN BY
 CC AN AMIDE BOND (BY SIMILARITY).
 DR EMBL; AB055806; BAB62042.1; -.
 DR InterPro; IPR005201; Glyco_hydro_85.
 DR InterPro; IPR005877; Gpos_YsIRK.
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR InterPro; IPR006192; LPXTG.
 DR Pfam; PF03644; Glyco_hydro_85; 1.
 DR Pfam; PF00746; Gram_pos_anchor; 1.
 DR Pfam; PF04650; YsIRK_signal; 1.
 DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
 DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
 DR TIGRFAMs; TIGR01168; YsIRK_signal; 1.
 DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
 KW Cell wall; Peptidoglycan-anchor.
 SQ SEQUENCE 1646 AA; 182121 MW; 5050E572B8E58D4 CRC64;

Query Match 1.0%; Score 8; DB 2; Length 1646;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 447 LTEAHKAL 454
 Db 1279 LTEAHKAL 1286

RESULT 46
 ID Q97S90 PRELIMINARY; PRT; 1659 AA.
 AC Q97S90;
 DT 01-OCT-2001 (TREMblrel. 18, Created)
 DT 01-OCT-2001 (TREMblrel. 18, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Endo-beta-N-acetylglucosaminidase, putative.
 GN SP0498.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TIGR4;
 RX MEDLINE=21357209; PubMed=11463916;
 RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
 RA Peterson S., Heidelberg J., DeBoy R.T., Halt D.H., Dodson R.J.,
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radue D.,
 RA Holtzapfel E., Khouli H., Wolf A.M., Uterback T.R., Hansen C.L.,
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
 RT "Complete genome sequence of a virulent isolate of *Streptococcus*
 RT pneumoniae.";
 RT Science 293:498-506(2001).
 RL EMBL; AE007361; AAK74556.1; -.
 DR TIGR; SP0498; -.
 DR InterPro; IPR005201; Glyco_hydro_85.
 DR InterPro; IPR005877; Gpos_YsIRK.
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR InterPro; IPR006192; LPXTG.
 DR Pfam; PF03644; Glyco_hydro_85; 1.
 DR Pfam; PF00746; Gram_pos_anchor; 1.
 DR Pfam; PF04650; YsIRK_signal; 1.
 DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
 DR TIGRFAMs; TIGR01168; YsIRK_signal; 1.
 DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
 KW Complete proteome.
 SQ SEQUENCE 1659 AA; 183202 MW; 06A52DACA96AB074 CRC64;

Query Match 1.0%; Score 8; DB 16; Length 1659;
 Best Local Similarity 100.0%; Pred. No. 3e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 447 LTEAHKAL 454
 Db 1292 LTEAHKAL 1299

RESULT 47
 ID Q8CZ52 PRELIMINARY; PRT; 1659 AA.
 AC Q8CZ52;
 DT 01-MAR-2003 (TREMblrel. 23, Created)
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Hypothetical protein.
 GN SP00440.
 OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=171101;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21429245; PubMed=11544234;

RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA DeHoff B.S., Estrem S.T., Filtz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Khoja H., Kratz A.R., Lagace R.E.,
RA Leblanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McArthur S.M., McHenry M., McMaster K., Mundy C.W., Niclas T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-W., Winkler W.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Roestek P.R. Jr., Skatrud P.L.,
RA Glaes J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
DR EMBL: AB008424; AAK9244.1; -;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1659 AA; 183107 MW; EF072B217E4AD32 CRC64;

Query Match 1.0%; Score 8; DB 16; Length 1659;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 447 LTEAHKAL 454
Db 1292 LTEAHKAL 1299

RESULT 48
Q8SSW5 PRELIMINARY; PRT; 2621 AA.
ID Q8SSW5
AC Q8SSW5;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-UN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE KIAA1058 protein.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
NCBI_Taxid=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MX4;
RA Glockner G., Eichinger L., Szafranski K., Pachbat J., Dear P.,
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC115585; AAL92252.1; -;
SQ SEQUENCE 2621 AA; 293740 MW; EF6A8B5C1B7D210 CRC64;

Query Match 1.0%; Score 8; DB 5; Length 2621;
Best Local Similarity 100.0%; Pred. No. 4.5e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 252 SNNSNTNS 259
Db 19 SNNSNTNS 26

RESULT 49
O87936 PRELIMINARY; PRT; 42 AA.
ID O87936
AC O87936;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ENV protein (Fragment).
GN ENV.
OS Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
NCBI_Taxid=11723;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=macC8;
RX MEDLINE=95333295; PubMed=7609080;
RA Whatmore A., Cook N., Hall G., Sharpe S., Rud E., Cranage M.;
RT "Repair and evolution of nef in vivo modulates simian immunodeficiency

RT virus virulence.";
RL J. Virol. 69:5117-5123(1995).
DR EMBL: X86726; CAA60398.1; -;
DR InterPro: IPR000328; Env_Gp41.
KW Pfam: PF00517; Gp41; 1.
KW Transmembrane.
FT NON TER
SQ SEQUENCE 42 AA; 4890 MW; F1E9ADBB76A83E7A CRC64;

Query Match 0.9%; Score 7; DB 15; Length 42;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 750 ATEYLAG 756
Db 4 ATEYLAG 10

RESULT 50
O87948 PRELIMINARY; PRT; 42 AA.
ID O87948
AC O87948;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ENV protein (Fragment).
GN ENV.
OS Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
NCBI_Taxid=11723;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=macC8;
RX MEDLINE=95333295; PubMed=7609080;
RA Whatmore A., Cook N., Hall G., Sharpe S., Rud E., Cranage M.;
RT "Repair and evolution of nef in vivo modulates simian immunodeficiency
virus virulence.";
RL J. Virol. 69:5117-5123(1995).
DR EMBL: X86731; CAA60408.1; -;
DR InterPro: IPR000328; Env_Gp41.
KW Pfam: PF00517; Gp41; 1.
KW Transmembrane.
FT NON TER
SQ SEQUENCE 42 AA; 4890 MW; F1E9ADBB76A83E7A CRC64;

Query Match 0.9%; Score 7; DB 15; Length 42;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 750 ATEYLAG 756
Db 4 ATEYLAG 10

RESULT 51
O87938 PRELIMINARY; PRT; 42 AA.
ID O87938
AC O87938;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ENV protein (Fragment).
GN ENV.
OS Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
NCBI_Taxid=11723;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=macC8;
RX MEDLINE=95333295; PubMed=7609080;
RA Whatmore A., Cook N., Hall G., Sharpe S., Rud E., Cranage M.;
RT "Repair and evolution of nef in vivo modulates simian immunodeficiency
virus virulence.";

RL J. Virol. 69:5117-5123(1995).
 DR EMBL: X86724; CAA60394.1; -;
 DR InterPro: IPR000328; Env_GP41.
 DR Pfam: PF00517; GP41; 1.
 KM Transmembrane.
 FT NON TER
 SQ SEQUENCE 42 AA; 4890 MW; F1E9ADB76A83E7A CRC64;

Query Match 0.9%; Score 7; DB 15; Length 42;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY 750 ATEFLAG 756
 |||||
 Db 4 ATEFLAG 10

RESULT 52
 Q88084 PRELIMINARY; PRT; 42 AA.
 ID Q88084
 AC Q88084
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE ENV protein (Fragment).
 GN ENV.
 OS Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11723;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=macC8;
 RA Whatmore A.M.;
 RL Submitted (Aug-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL: X90853; CAA62360.1; -;
 DR InterPro: IPR000328; Env_GP41.
 DR Pfam: PF00517; GP41; 1.
 KM Transmembrane.
 FT NON TER
 SQ SEQUENCE 42 AA; 4871 MW; F1F57C7676A83E7A CRC64;
 Query Match 0.9%; Score 7; DB 15; Length 42;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY 750 ATEFLAG 756
 |||||
 Db 4 ATEFLAG 10

RESULT 53
 Q87952 PRELIMINARY; PRT; 42 AA.
 ID Q87952
 AC Q87952
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE ENV protein (Fragment).
 GN ENV.
 OS Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11723;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=macC8;
 RA Whatmore A.M.;
 RL Submitted (Aug-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL: X86725; CAA60396.1; -;
 DR InterPro: IPR000328; Env_GP41.
 DR Pfam: PF00517; GP41; 1.

KM Transmembrane.
 FT NON TER
 SQ SEQUENCE 42 AA; 4890 MW; F1E9ADB76A83E7A CRC64;
 Query Match 0.9%; Score 7; DB 15; Length 42;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY 750 ATEFLAG 756
 |||||
 Db 4 ATEFLAG 10

RESULT 54
 Q87946 PRELIMINARY; PRT; 42 AA.
 ID Q87946
 AC Q87946
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE ENV protein (Fragment).
 GN ENV.
 OS Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11723;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=macC8;
 RA Whatmore A.M.;
 RL Submitted (Aug-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL: X86730; CAA60406.1; -;
 DR InterPro: IPR000328; Env_GP41.
 DR Pfam: PF00517; GP41; 1.
 KM Transmembrane.
 FT NON TER
 SQ SEQUENCE 42 AA; 4890 MW; F1E9ADB76A83E7A CRC64;
 Query Match 0.9%; Score 7; DB 15; Length 42;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY 750 ATEFLAG 756
 |||||
 Db 4 ATEFLAG 10

RESULT 55
 Q88080 PRELIMINARY; PRT; 42 AA.
 ID Q88080
 AC Q88080
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE ENV protein (Fragment).
 GN ENV.
 OS Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11723;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=macC8;
 RA Whatmore A.M.;
 RL Submitted (Aug-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL: X90851; CAA62356.1; -;
 DR InterPro: IPR000328; Env_GP41.
 DR Pfam: PF00517; GP41; 1.
 KM Transmembrane.
 FT NON TER
 SQ SEQUENCE 42 AA; 4818 MW; F1EB8D9976A83E7A CRC64;

Query Match 0.9%; Score 7; DB 15; Length 42;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 750 ATETLAG 756
|||||
Db 4 ATETLAG 10

RESULT 56
087944 PRELIMINARY; PRT; 43 AA.
AC 087944;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE ENV protein (Fragment).
GN ENV.
OS Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11723;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=macC8;
RX MEDLINE=95333295; PubMed=7609080;
RA Whatmore A., Cook N., Hall G., Sharpe S., Rud E., Cranage M.;
RT "Repair and evolution of nef in vivo modulates simian immunodeficiency virus virulence".
RL J. Virol. 69:5117-5123(1995).
DR EMBL; X86729; CAA60404.1; -;
DR InterPro; IPR000328; Env_GP41.
DR Pfam; PF00517; GP41; 1.
KW Transmembrane.
FT NON_TER 1
SQ SEQUENCE 43 AA; 4947 MW; F1E9FF5CE833E7A CRC64;

Query Match 0.9%; Score 7; DB 15; Length 43;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 750 ATETLAG 756
|||||
Db 5 ATETLAG 11

RESULT 57
087942 PRELIMINARY; PRT; 44 AA.
ID 087942;
AC 087942;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE ENV protein (Fragment).
GN ENV.
OS Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11723;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=macC8;
RX MEDLINE=95333295; PubMed=7609080;
RA Whatmore A., Cook N., Hall G., Sharpe S., Rud E., Cranage M.;
RT "Repair and evolution of nef in vivo modulates simian immunodeficiency virus virulence".
RL J. Virol. 69:5117-5123(1995).
DR EMBL; X86728; CAA60402.1; -;
DR InterPro; IPR000328; Env_GP41.
DR Pfam; PF00517; GP41; 1.
KW Transmembrane.
FT NON_TER 1
SQ SEQUENCE 44 AA; 5046 MW; F1E9FF1A4A09507A CRC64;

Query Match 0.9%; Score 7; DB 15; Length 44;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 750 ATETLAG 756
|||||
Db 6 ATETLAG 12

RESULT 58
087940 PRELIMINARY; PRT; 49 AA.
ID 087940;
AC 087940;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE ENV protein (Fragment).
GN ENV.
OS Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11723;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=macC8;
RX MEDLINE=95333295; PubMed=7609080;
RA Whatmore A., Cook N., Hall G., Sharpe S., Rud E., Cranage M.;
RT "Repair and evolution of nef in vivo modulates simian immunodeficiency virus virulence".
RL J. Virol. 69:5117-5123(1995).
DR EMBL; X86727; CAA60400.1; -;
DR InterPro; IPR000328; Env_GP41.
DR Pfam; PF00517; GP41; 1.
KW Transmembrane.
FT NON_TER 1
SQ SEQUENCE 49 AA; 5601 MW; A4C9F81D7A5BF6A5 CRC64;

Query Match 0.9%; Score 7; DB 15; Length 49;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 750 ATETLAG 756
|||||
Db 11 ATETLAG 17

RESULT 59
Q45596 PRELIMINARY; PRT; 49 AA.
ID Q45596;
AC Q45596;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Function unknown.
GN YYDF.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96051385; PubMed=7584024;
RA Ogasawara N., Nakai S., Yoshikawa H.;
RT "Systematic sequencing of the 180 kilobase region of the Bacillus subtilis chromosome containing the replication origin".
RL DNA Res. 1:1-14(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=94171085; PubMed=8125345;
RA Zhang J., Aronson A.T.;
RT "A Bacillus subtilis bglA gene encoding phospho-beta-glucosidase is inducible and closely linked to a NADH dehydrogenase-encoding gene".
RL Gene 140:85-90(1994).
RN [3]

RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=94156824; PubMed=8113162;
 RA Calogero S., Gardan R., Glaser P., Schweizer J., Rapoport G.,
 RT Debarbouille M.;
 RT "RocR, a novel regulatory protein controlling arginine utilization in
 RT Bacillus subtilis, belongs to the NtrC/NIFA family of transcriptional
 RT activators.";
 RL J. Bacteriol. 176:1234-1241 (1994).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=94236234; PubMed=8180695;
 RA Hartford O.M., Dowds B.C.;
 RT "Isolation and characterization of a hydrogen peroxide resistant
 RT mutant of Bacillus subtilis";
 RL Microbiology 140:297-304 (1994).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=96093926; PubMed=7584049;
 RA Yoshida K., Seki S., Fujimura M., Miwa Y., Fujita Y.;
 RT "Cloning and sequencing of a 36-Kb region of the Bacillus subtilis
 RT genome between the gnt and tol operons.";
 RL DNA Res. 2:61-69 (1995).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=95311309; PubMed=7540694;
 RA Gardan R., Rapoport G., Debarbouille M.;
 RT "Expression of the rocDEF operon involved in arginine catabolism in
 RT Bacillus subtilis.";
 RL J. Mol. Biol. 249:843-856 (1995).
 RN [7]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX Kaahara Y., Nakai S., Yoshikawa H., Ogasawara N.;
 RL Submitted (NOV-1995) to the EMBL/Genbank/DBJ databases.
 RN [8]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertoro M.G., Bessieres P., Bolotin A., Borcherst S.,
 RA Borriss R., Boursier L., Brans A., Braun M., Brigelli S.C., Bron S.,
 RA Brouillet S., Brunsch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Conteron I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entian K.D., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Golligly E.J., Grandi G.,
 RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Katamata D., Kasahara Y., Klaerr-Bianchard M., Klein C.,
 RA Kobayashi Y., Koeltter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Priesemann E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeder R., Scoffone F.,
 RA Setiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
 RA Tokokin A., Tacconi H., Takagi T., Takahashi H., Takenaru K.,
 RA Takeuchi M., Tatemochi A., Tanaka T., Terpsira P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambuit R., Wedler E., Wedler H., Wellenreger T.,
 RA Witters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zunschein E., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the gram-positive bacterium Bacillus
 RT subtilis.";
 RL Nature 390:249-256 (1997).

RN [9]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
 RL Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.
 DR EMBL; D78193; BA01276.1; -;
 DR EMBL; Z99124; CAB16055.1; -;
 RN Complete proteome.
 SQ SEQUENCE 49 AA; 5781 MW; BA5F90DAD2982A86 CRC64;
 QY 406 ETVKNLE 412
 DB 9 ETVKNLE 15
 Query Match 0.9%; Score 7; DB 16; Length 49;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 60
 QY 406 ETVKNLE 412
 DB 9 ETVKNLE 15
 Query Match 0.9%; Score 7; DB 16; Length 49;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 61
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 62
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 63
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 64
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 65
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 66
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 67
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 68
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 69
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 70
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 71
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 72
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 73
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 74
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 75
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 76
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 77
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 78
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 79
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 80
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 81
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 82
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 83
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 84
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 85
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 86
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 87
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 88
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 89
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 90
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 91
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 92
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 93
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 94
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 95
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 96
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 97
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 98
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 99
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 100
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 101
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 102
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 103
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 104
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 105
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 106
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 107
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 108
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 109
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 110
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 111
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 112
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 113
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 114
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 115
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 116
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 117
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 118
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 119
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 120
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 121
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 122
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 123
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 124
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 125
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 126
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 127
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 128
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 129
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 130
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 131
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 132
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 133
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 134
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 135
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 136
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 137
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 138
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 139
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 140
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 141
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 142
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 143
 QY 97 IYNEVKG 103
 DB 25 IYNEVKG 31
 Query Match 0.9%; Score 7; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 144
 QY 97 I

RC STRAIN=06:HI / CFT073 / ATCC 700928;
RA MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G., III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Domeneberg M.S., Blattner F.R.,
RT "Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic Escherichia coli.",
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AE016760; AAN80206.1; -;
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 51 AA; 603 MW; 7BCD4D2E0A509E0B CRC64;

QY Query Match 0.9%; Score 7; DB 16; Length 51;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 520 LADKYTT 526
DB 18 LADKYTT 24

RESULT 62

Q90VA0 PRELIMINARY; PRT; 68 AA.
AC Q90VA0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Nef protein (Negative factor) (27 kDa protein) (Fragment).
GN NEF.
OS Human immunodeficiency virus 1.
OC Viruses; Retroviruses; Retroviridae; Lentivirus.
OX NCBI_Taxid=11676;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B+48;
RX MEDLINE=21160457; PubMed=11259184;
RA Ibanez A., Cloet B., Martinez M.A.;
RT "Absence of genetic diversity reduction in the HIV-1 integrated
RT proviral LTR sequence population during successful combination
RT therapy.",
RL Virology. 282:1-5(2001).
CC -1- FUNCTION: NEF HAS GTPASE, GTP-BINDING AND AUTOPHOSPHORYLATING
CC ACTIVITIES. IT SEEMS TO DOWN-REGULATE THE CD4 (T4) ANTIGEN (BY
CC SIMILARITY).
DR EMBL; AF294070; AAK48560.1; -;
DR InterPro; IPR001558; HIV_Nef.
DR Pfam; PF00469; F-Protein; 1.
DR ProDom; PD000031; HIV_Nef; 1.
DR AIDS; GTP-binding; lipidprotein; Myristate.
FT NON_TER 1
SQ SEQUENCE 68 AA; 8020 MW; F10EC7AB1A8D247F CRC64;

QY Query Match 0.9%; Score 7; DB 15; Length 68;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 722 VETEKVE 728
DB 10 VETEKVE 16

RESULT 63

Q9A3B0 PRELIMINARY; PRT; 68 AA.
AC Q9A3B0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein CC3294.
GN CC3294.
OS Caulobacter crescentum.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;
OC Caulobacteraceae; Caulobacter.
OX NCBI_Taxid=155892;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentum.",
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005992; AAK25256.1; -;
DR TIGR; CC3294; -;
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 68 AA; 7458 MW; FAB37837A43B2E82 CRC64;

QY Query Match 0.9%; Score 7; DB 16; Length 68;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 322 ELEERIA 328
DB 31 ELEERIA 37

RESULT 64

Q977L2 PRELIMINARY; PRT; 71 AA.
AC Q977L2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Zn-ribbon protein.
OS uncultured crenarchaeote 7444.
OC Archaea; Crenarchaeota; environmental samples;
OC marine archaeal group 1.
OX NCBI_Taxid=166279;
RN [1]
RP SEQUENCE FROM N.A.
RA Beja O., Koonin E.V., Aravind L., Taylor L.T., Seiz H., Stein J.L.,
RA Bensen D.C., Feldman R.A., Swanson R.V., DeLong E.F.;
RT "Comparative genomic analysis of coexisting archaeal genetic variants
RT in an Antarctic marine microbial assemblage.",
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF393466; AAK96096.1; -;
SQ SEQUENCE 71 AA; 8052 MW; C14F0FB3B9858FE CRC64;

QY Query Match 0.9%; Score 7; DB 1; Length 71;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 710 VETPAE 716
DB 16 VETPAE 22

RESULT 65

Q96YD8 PRELIMINARY; PRT; 71 AA.
AC Q96YD8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Putative integrase.
GN STS236.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.

OX NCBI_TaxID=111955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JCM 10545 / 7;
 RX PubMed=11572479;
 RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
 RA Sekine M., Baba S.-I., Ankaï H., Kosugi H., Hosoyama A., Fukui S.,
 RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
 RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
 RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
 RA Oshima T., Kikuchi H.;
 RT "Complete genome sequence of an aerobic thermocacidophilic
 RT Crenarchaeon, *Sulfolobus tokodaii* strain7.";
 RL DNA Res. 8:123-140(2001).
 DR EMBL: AP000989; BAB67339.1; -;
 KW Hypothetical protein; Complete proteome.
 SO SEQUENCE 71 AA; 8115 MW; 3D46F83EBB546E37 CRC64;

Query Match 0.9%; Score 7; DB 17; Length 71;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 GKYYVYL 117
 Db 18 GKYYVYL 24

RESULT 66

ID Q9HGR7 PRELIMINARY; PRT; 80 AA.
 AC Q9HGR7;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12) (GAPDH)
 DE (Fragment).
 GN GPD.
 OS *Gilbertella periscaria*.
 OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Gilbertellaceae;
 OC Gilbertella.
 OX NCBI_TaxID=101096;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tamas P.;
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 RP [2]
 RP SEQUENCE FROM N.A.
 RA Papp T., Vastag M., Acs K., Vagvolgyi C.;
 RT "Phylogenetic relationships among Mucoraceae, Choanephoraceae and
 RT Gilbertellaceae based on rDNA and glyceraldehyde-3-phosphate
 RT dehydrogenase sequences.";
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + PHOSPHATE +
 CC NAD(+) = 3-PHOSPHO-D-GLYCEROL PHOSPHATE + NADH.
 CC -1- PATHWAY: SECOND PHASE OF GLYCOLYSIS; FIRST STEP.
 CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY)
 CC -1- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE
 CC DEHYDROGENASE FAMILY.
 DR EMBL: AJ278316; CAC05663.1; -;
 DR HSSP; P00354; 3GPD.
 DR InterPro: IPR001173; GAP_dhhdhgenase.
 DR Pfam; PF02800; gpdh_C; 1.
 KW Glycolysis; NAD; Oxidoreductase.
 FT NON_TER 1 1
 FT NON_TER 80 80
 SO SEQUENCE 80 AA; 8609 MW; A571C27B7C915F0E CRC64;

Query Match 0.9%; Score 7; DB 3; Length 80;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 509 EYTEDEV 515

Db 43 EYTEDEV 49
 |||||

RESULT 67

ID Q8MYJ9 PRELIMINARY; PRT; 83 AA.
 AC Q8MYJ9;
 DT 01-OCT-2002 (TREMBlrel. 22, Created)
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE U6 snRNA-associated Sm-like protein Lsm6.
 DE Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX4;
 RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
 RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
 RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
 RT "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC116961; AAM3742.1; -;
 DR InterPro: IPR006649; snRNP.
 DR InterPro: IPR001163; snRNP_Sm.
 DR Pfam; PF01423; Sm; 1.
 DR ProDom; PD020287; snRNP; 1.
 DR SMART; SM00551; Sm; 1.
 SO SEQUENCE 83 AA; 9411 MW; 8299A18276894BEA CRC64;

Query Match 0.9%; Score 7; DB 5; Length 83;
 Best Local Similarity 100.0%; Pred. No. 2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 250 NTSNNSN 256
 Db 9 NTSNNSN 15

RESULT 68

ID Q73850 PRELIMINARY; PRT; 86 AA.
 AC Q73850;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Envelope glycoprotein, V1-V2 region (Fragment).
 GN ENV.
 OS Human immunodeficiency virus 1.
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11676;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96251940; PubMed=8661394;
 RA Palmer C., Balfe P., Fox D., May J.C., Frederiksen R., Fenyo E.M.,
 RA McKeating J.A.;
 RT "Functional characterization of the V1V2 region of human
 RT immunodeficiency virus type 1.";
 RL Virology 220:436-449(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Blouin C.;
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U52268; AAB07922.1; -;
 DR InterPro: IPR000777; GP120.
 DR Pfam; PF00516; GP120; 1.
 KW AIDS; Coat protein; Glycoprotein.
 FT NON_TER 1 1
 FT NON_TER 86 86
 SO SEQUENCE 86 AA; 9788 MW; 6241F1B599434EF3 CRC64;

Query Match 0.9%; Score 7; DB 15; Length 86;

Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 247 TNTNTSN 253
Db 25 TNTNTSN 31

RESULT 69
073851

ID 073851 PRELIMINARY; PRT; 86 AA.

AC 073851; 01-NOV-1996 (TReMBLrel. 01, Created)

DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)

DE 01-OCT-2002 (TReMBLrel. 22, Last annotation update)

ENV Envelope glycoprotein, V1-V2 region (Fragment).

OS Human immunodeficiency virus 1.

OC Viruses; Retrovird viruses; Retroviridae; Lentivirus.

OX NCBI_TaxID=11676;

RN [1] SEQUENCE FROM N.A.

RA MEDLINE=96251940; PubMed=8661394;

RT Palmer C., Baile P., Fox D., May J.C., Frederiksson R., Fenyo E.M.,

MCKeating J.A.;

RT "Functional characterization of the V1V2 region of human

immunodeficiency virus type 1."

RL Virology 220:436-449(1996).

RN [2] SEQUENCE FROM N.A.

RA Blouin C.;

RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL; US2269; AAB07923.1; -

DR Interpro: IPR000777; GP120.

DR Pfam: PF00516; GP120; 1.

KW AIDS; Coat protein; Glycoprotein.

FT NON_TER 1 1

FT NON_TER 86 86

FT NON_TER 9788 9788

SO SEQUENCE 86 AA; 9788 MW; 6241F1E599434EF3 CRC64;

Query Match 0.9%; Score 7; DB 15; Length 86;

Best Local Similarity 100.0%; Pred. No. 2e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 247 TNTNTSN 253

Db 25 TNTNTSN 31

RESULT 70
073852

ID 073852 PRELIMINARY; PRT; 86 AA.

AC 073852; 01-NOV-1996 (TReMBLrel. 01, Created)

DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)

DE 01-OCT-2002 (TReMBLrel. 22, Last annotation update)

ENV Envelope glycoprotein, V1-V2 region (Fragment).

OS Human immunodeficiency virus 1.

OC Viruses; Retrovird viruses; Retroviridae; Lentivirus.

OX NCBI_TaxID=11676;

RN [1] SEQUENCE FROM N.A.

RA MEDLINE=96251940; PubMed=8661394;

RT Palmer C., Baile P., Fox D., May J.C., Frederiksson R., Fenyo E.M.,

MCKeating J.A.;

RT "Functional characterization of the V1V2 region of human

immunodeficiency virus type 1."

RL Virology 220:436-449(1996).

RN [2] SEQUENCE FROM N.A.

RA Blouin C.;

RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL; US2270; AAB07924.1; -

DR Interpro: IPR000777; GP120.

DR Pfam: PF00516; GP120; 1.

KW AIDS; Coat protein; Glycoprotein.

FT NON_TER 1 1

FT NON_TER 86 86

FT NON_TER 9788 9788

SO SEQUENCE 86 AA; 9788 MW; 6241F1E599434EF3 CRC64;

Query Match 0.9%; Score 7; DB 15; Length 86;

Best Local Similarity 100.0%; Pred. No. 2e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 247 TNTNTSN 253

Db 25 TNTNTSN 31

RESULT 71
097W98

ID 097W98 PRELIMINARY; PRT; 90 AA.

AC 097W98; 01-OCT-2001 (TReMBLrel. 18, Created)

DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)

DE 01-OCT-2001 (TReMBLrel. 18, Last annotation update)

ENV Hypothetical protein SSO10237.

OS Sulfolobus solfataricus.

OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;

OX NCBI_TaxID=2287;

RN [1] SEQUENCE FROM N.A.

RA STRAIN=ATCC 35092 / DSM 1617 / P2;

RA MEDLINE=21332296; PubMed=11427726;

RA She Q., Singh R.K., Contaloni F., Zivanovic Y., Allard G.,

Awatey M.J., Chan-Welher C.C.-Y., Clausen I.G., Cutlis B.A.,

De Moers A., Erauso G., Fletcher C., Gordon P.M.K.,

RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,

RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,

RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,

RT Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;

RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";

RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).

DR EMBL; AE006836; AAK42490.1; -

KW Hypothetical protein; Complete proteome.

SO SEQUENCE 90 AA; 10208 MW; 5AE75F77EFED0635 CRC64;

Query Match 0.9%; Score 7; DB 17; Length 90;

Best Local Similarity 100.0%; Pred. No. 2.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 GKXYVYL 117

Db 18 GKXYVYL 24

RESULT 72
09RUG5

ID 09RUG5 PRELIMINARY; PRT; 93 AA.

AC 09RUG5; 01-MAY-2000 (TReMBLrel. 13, Created)

DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)

DE 01-MAR-2002 (TReMBLrel. 20, Last annotation update)

ENV Hypothetical protein DR1421.

OS Deinococcus radiodurans.

OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;

OC Deinococcaceae; Deinococcus.

OX NCBI_TaxID=1299;

RN [1] SEQUENCE FROM N.A.

RA STRAIN=R1;

RX MEDLINE=20036896; PubMed=10567266;

```

RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson M.C., Richardson D.L.,
RA McFarr K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Uetreack T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577 (1999).
DR EMBL; AE001987; AAF11000.1; -.
DR TIGR; DR1421; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 93 AA; 10167 MW; 39DBF6F8DE24B5 CRC64;

Query Match 0.9%; Score 7; DB 16; Length 93;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 487 DLAFIA 493
Db 6 DLAFIA 12

RESULT 73
QRY27 PRELIMINARY; PRT; 94 AA.
AC Q9RY27;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein DR0124.
GN DR0124.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson M.C., Richardson D.L.,
RA McFarr K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Uetreack T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577 (1999).
DR EMBL; AE001875; AAF09721.1; -.
DR TIGR; DR0124; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 94 AA; 11353 MW; 4DFCB8011DE91235 CRC64;

Query Match 0.9%; Score 7; DB 16; Length 94;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 322 ELERIA 328
Db 81 ELERIA 87

RESULT 74
Q8FAP6 PRELIMINARY; PRT; 96 AA.
AC Q8FAP6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Transposase incs for insertion element IS2A/D/F/H/I/K.

```

```

GN C5196.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CPT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liu S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
DR EMBL; AE016771; AAN83618.1; -.
KW Complete proteome.
SQ SEQUENCE 96 AA; 10661 MW; 362073321AC7BD1E CRC64;

Query Match 0.9%; Score 7; DB 16; Length 96;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 202 ASEIAA 208
Db 45 ASEIAA 51

RESULT 75
Q8U029 PRELIMINARY; PRT; 104 AA.
AC Q8U029;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Small nucleolar rnp similar to gar1.
GN PF1791.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE010277; AAL81915.1; -.
KW Complete proteome.
SQ SEQUENCE 104 AA; 12311 MW; EFOA25A67B14687C CRC64;

Query Match 0.9%; Score 7; DB 17; Length 104;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 RTNWVS 240
Db 26 RTNWVS 32

RESULT 76
P88557 PRELIMINARY; PRT; 108 AA.
AC P88557;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE P17 (Gag polyprotein) (Fragment).
GN GAG.
OS Human immunodeficiency virus 1.
OC Viruses; Retrod viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97151115; PubMed=8995651;
RT Hughes E.S., Bell J.E., Simmonds P.;
RT "Investigation of the dynamics of the spread of human immunodeficiency
RT virus to brain and other tissues by evolutionary analysis of sequences
RT from the p17ag and env genes."
RL J. Virol. 71:1272-1280(1997).
DR EMBL; U79816; AAB46537.1; -
DR InterPro; IPR000071; Retrovir_p17.
DR Pfam; PF00540; Gag_p17.1.
DR PRINTS; PR00234; HIVMATRIX.
KM AIDS; Core Protein; PolyProtein.
FT NON_TER 1
FT NON_TER 108
SQ SEQUENCE 108 AA; 12031 MW; 06C862CF21B76E42 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 15; Length 108;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 372 NSLSVSG 378
Db 85 NSLSVSG 91

RESULT 77
08VR72 PRELIMINARY; PRT; 109 AA.
AC 08VR72;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Insertion sequence protein.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CF703;
RX MEDLINE=21458820; PubMed=11574920;
RX Rasio D.A., Phillips J.A., Li X., Mobley H.L.T.;
RT "Identification of DNA sequences from a second pathogenicity island of
RT uropathogenic Escherichia coli CFT073: probes specific for
RT uropathogenic populations."
RL J. Infect. Dis. 184:1041-1049(2001).
DR EMBL; AP447814; AAL67376.1; -
DR InterPro; IPR002514; Transposase_8.
DR Pfam; PF01527; Transposase_8; 1.
SQ SEQUENCE 109 AA; 12173 MW; 5546C66A2BB96F28 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 2; Length 109;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 202 ASELIANA 208
Db 70 ASELIANA 76

RESULT 78
09A6S4 PRELIMINARY; PRT; 110 AA.
AC 09A6S4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hsb/Yadr/Yfhf family protein.
DE CC2009.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;
OC Caulobacteraceae; Caulobacter.

```

```

OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RX Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RX Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RX Potocka I., Nelson W.C., Newton A., Stephens C., Padde N.D., Ely B.,
RX DeBoy R.T., Dodson R.J., Durkin A.S., Gunn M.L., Haft D.H.,
RX Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RX Uetback T., Tran K., Wolf A., Yamanevan U., Ermolaeva M., White O.,
RX Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005874; AAK23984.1; -
DR TIGR; CC2009; -
DR InterPro; IPR000361; Hsb_Yadr_Yfhf.
DR Pfam; PF01521; Hsb-1like; 1.
DR ProDom; PD002183; Hsb_Yadr_Yfhf; 1.
DR TIGRPFAM; TIGR00049; TIGR00049; 1.
DR PROSITE; PS01152; HSB; 1.
KM Complete proteome.
SQ SEQUENCE 110 AA; 11698 MW; D1DCD55226D9119 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 16; Length 110;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 779 LALLKGS 785
Db 72 LALLKGS 78

RESULT 79
09CDT0 PRELIMINARY; PRT; 110 AA.
AC 09CDT0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Unknown protein.
GN YWAB OR IL2131.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RX Bolotin A., Wincker P., Manger S., Jallion O., Malarne K.,
RX Weisenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403."
RL Genome Res. 11:731-753(2001).
DR EMBL; AE006441; AAK06229.1; -
DR InterPro; IPR001412; tRNA-synt_1.
DR Pfam; PF02699; YajC; 1.
DR TIGRPFAM; TIGR00739; YajC; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KM Complete proteome.
SQ SEQUENCE 110 AA; 11915 MW; FA8E3A9C84DAE2CE CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 16; Length 110;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 710 VEETPAE 716
Db 94 VEETPAE 100

RESULT 80

```

08FC10 PRELIMINARY; PRT; 111 AA.
 ID 08FC10: 01-MAR-2003 (TREMBlrel. 23, Created)
 AC 08FC10: 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DE Transposase *insC* for insertion element IS2A/D/F/H/I/K.
 GN C4552.
 OS *Escherichia coli* O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=217992;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O6:H1 / ATCC 790928;
 RX MEDLINE=22388234; PubMed=12471157;
 RA Welch R.A., Birtland V., Plunkett G., III, Redford P., Roesch P.,
 RA Raeko D., Buckles E.L., Lion S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 of uropathogenic *Escherichia coli*."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
 DR EMBL; AE016769; AAN82986.1; -.
 KW Complete proteome.
 SQ SEQUENCE 111 AA; 12410 MW; 9A637F9998E413CA CRC64;

Query Match 0.9%; Score 7; DB 16; Length 111;
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 202 ASELAAA 208
 DB 85 ASELAAA 91

RESULT 81
 08CMN4 PRELIMINARY; PRT; 115 AA.
 ID 08CMN4: 01-MAR-2003 (TREMBlrel. 23, Created)
 AC 08CMN4: 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DE Ribonuclease P protein component.
 GN SE2418.
 OS *Staphylococcus epidermidis*.
 CC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=1282;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 12228;
 RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
 RA Chen Z., Wen Y.;
 RT Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE016752; AAO06061.1; -.
 KW Complete proteome.
 SQ SEQUENCE 115 AA; 13484 MW; 106B2592C400F18 CRC64;

Query Match 0.9%; Score 7; DB 16; Length 115;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 460 RNSDFQA 466
 DB 9 RNSDFQA 15

RESULT 82
 08T4Z7 PRELIMINARY; PRT; 116 AA.
 ID 08T4Z7: 01-JUN-2002 (TREMBlrel. 21, Created)
 AC 08T4Z7: 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Erythrocyte membrane protein 1 (Fragment).
 GN VAR.
 OS *Plasmodium falciparum*.
 CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PH1;
 RX MEDLINE=21839615; PubMed=11849711;
 RA Fowler E.V., Peters J.M., Gatton M.L., Chen N., Cheng Q.;
 RT "Genetic diversity of the DBLalpha region in *Plasmodium falciparum* var
 genes among Asia-Pacific isolates."
 RL Mol. Biochem. Parasitol. 120:117-126(2002).
 DR EMBL; AY054913; AAL11277.1; -.
 FT NON-TER
 FT NON-TER
 SQ SEQUENCE 116 AA; 13272 MW; 01F917589E19E327 CRC64;

Query Match 0.9%; Score 7; DB 5; Length 116;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 649 TYKAPNG 655
 DB 71 TYKAPNG 77

RESULT 83
 08L494 PRELIMINARY; PRT; 117 AA.
 ID 08L494: 01-OCT-2002 (TREMBlrel. 22, Created)
 AC 08L494: 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE OJ1316_H05.2 protein (P0038D11.19 protein).
 GN OJ1316_H05.2 OR P0038D11.19.
 OS *Oryza sativa* (japonica cultivar-group).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 CC Eriocarpaceae; Oryzae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, BAC
 clone:OJ1316_H05.";
 RT Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
 clone:P0038D11.19";
 RT Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP003340; BAC00607.1; -.
 DR EMBL; AP003234; BAC0543.1; -.
 DR Gramene; 08L494; -.
 SQ SEQUENCE 117 AA; 13260 MW; A03BF65BF4FE3194 CRC64;

Query Match 0.9%; Score 7; DB 10; Length 117;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 381 RKGEGY 387
 DB 83 RKGEGY 89

RESULT 84
 08YC21 PRELIMINARY; PRT; 117 AA.
 ID 08YC21

AC 08YC21;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, last annotation update)
 DE Transcriptional regulator.
 GN BMEI10712.
 OS Brucella melitensis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Brucella.
 OX NCBI_TaxID=29459;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=16M / ATCC 23456 / Biotype 1;
 RX MEDLINE=20020109; PubMed=11756688;
 RA DelVecchio V.G., Kapral V., Redkar R.J., Patra G., Mijer C., Los T.,
 Ivanova N., Anderson I., Bhattacharya A., Lykdis A., Resnik G.,
 RA Jahnke L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,
 Selkov E., Elzer F.H., Hegins S., O'Callaghan D., Letesson J.-J.,
 RA Haselkorn R., Kyriades N., Overbeek R.;
 RT "The genome sequence of the facultative intracellular pathogen
 RT Brucella melitensis."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
 DR EMBL; AE009706; AAL53954.1; -.
 KW Complete proteome.
 SQ SEQUENCE 117 AA; 12905 MW; B180B0A08BD6D49A CRC64;

Query Match 0.9%; Score 7; DB 16; Length 117;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 200 LSASELA 206
 Db 88 LSASELA 94

RESULT 85
 Q8OUR7 PRELIMINARY; PRT; 120 AA.
 AC 08OUR7;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, last annotation update)
 DE ORF043L.
 OS Infectious spleen and kidney necrosis virus.
 OC Viruses; dsDNA viruses, no RNA stage; Iridoviridae;
 OC Unclassified Iridoviridae.
 OX NCBI_TaxID=180170;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21874810; PubMed=11878882;
 RA He J.G., Deng M., Weng S.P., Li Z., Zhou S.Y., Long Q.X., Wang X.Z.,
 RA Chan S.M.;
 RT "Complete genome analysis of the mandarin fish infectious spleen and
 RT kidney necrosis Iridovirus."
 RL Virology 291:126-139(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA He J.G., Deng M., Weng S.P., Li Z., Zhou S.Y., Long Q.X., Wang X.Z.,
 RA Chan S.M.;
 RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF371960; AAL98767.1; -.
 DR InterPro; IPR006863; Evtl.Alt.
 DR Pfam; PF04777; Evtl.Alt; I.
 SQ SEQUENCE 120 AA; 13326 MW; A06BFFDB695CB787 CRC64;

Query Match 0.9%; Score 7; DB 12; Length 120;
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 300 TARGAV 306
 Db 110 TARGAV 116

RESULT 86
 Q8FDX0 PRELIMINARY; PRT; 124 AA.
 AC 08FDX0;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, last annotation update)
 DE Transposase inc for insertion element IS2A/D/F/H/I/K.
 GN C3612.
 OS Escherichia coli O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=217992;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=06:HI / CFT073 / ATCC 700928;
 RX MEDLINE=22388234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 RT of uropathogenic Escherichia coli."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
 DR EMBL; AE016766; AAN82060.1; -.
 KW Complete proteome.
 SQ SEQUENCE 124 AA; 13998 MW; F038388F93D92689 CRC64;

Query Match 0.9%; Score 7; DB 16; Length 124;
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 202 ASELA 208
 Db 85 ASELA 91

RESULT 87
 Q9X6J3 PRELIMINARY; PRT; 126 AA.
 AC 09X6J3;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, last annotation update)
 DE Chorismate mutase.
 GN AROH.
 OS Bacillus stearothermophilus.
 OC Bacteria; Firmicutes; Bacillales; Geobacillus.
 OX NCBI_TaxID=1422;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCA 26;
 RA Chen X.P., Antson A.A., Yang M., Baumann C., Dodson E.J., Dodson G.G.,
 RA Gollnick P.;
 RT "Regulatory features of the trp operon and the crystal structure of
 RT the trp RNA-binding Attenuation Protein from Bacillus
 RT stearothermophilus."
 RL J. Mol. Biol. 0:0-0(1999).
 DR EMBL; AF139534; AAD33790.1; -.
 DR HSSP; P19080; IDBF.
 DR InterPro; IPR002701; Chorismate_mut.
 DR Pfam; PF01817; Chorismate_mut; I.
 SQ SEQUENCE 126 AA; 14101 MW; 0761341DF8BD7CB4 CRC64;

Query Match 0.9%; Score 7; DB 2; Length 126;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 115 VYLKDA 121
 Db 106 VYLKDA 112

RESULT 88

Q8FWA2 PRELIMINARY; PRT; 128 AA.
 AC Q8FWA2;
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE 1566 family element, orf2, putative.
 GN BRA0557.
 OS Brucella suis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Brucella.
 OX NCBI_TaxID=29461;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1330 / Biovar 1;
 RC MEDLINE=22247741; PubMed=12271122;
 RX Paulsen I.T., Seehardt R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
 RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beaman M.J.,
 RA Daugherty S.C., Debey R.T., Durkin A.S., Kolonay J.F., Madupu R.,
 RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek U., Van Aken S.E.,
 RA Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
 RA Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Fraser C.M.,
 RT "The Brucella suis genome reveals fundamental similarities between
 RT animal and plant pathogens and symbionts."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
 DR EMBL; AE014552; J05471.1;
 DR TIGR; BRA0557;
 KW Complete proteome.
 SQ SEQUENCE 128 AA; 13913 MW; B702E85108AB0AA CRC64;

Query Match 0.9%; Score 7; DB 16; Length 128;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 200 LSASELA 206
 |||||
 Db 99 LSASELA 105

RESULT 89

Q8EMH8 PRELIMINARY; PRT; 128 AA.
 AC Q8EMH8;
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE 5-carboxymethyl-2-hydroxymuconate isomerase (EC 5.3.3.10).
 GN O82864.
 OS Oceanobacillus thelyensis.
 OC Bacteria; Firmicutes; Bacillales; Oceanobacillus.
 OX NCBI_TaxID=182710;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HTE831 / DSM 14371 / JCM 11309;
 RC MEDLINE=22220767; PubMed=12235376;
 RX Takami H., Takaki Y., Uchiyama I.,
 RA "Genome sequence of Oceanobacillus thelyensis isolated from the Iheya
 RT Ridge and its unexpected adaptive capabilities to extreme
 RT environments."
 RL Nucleic Acids Res. 30:3927-3935(2002).
 DR EMBL; AP004602; BAC14820.1;
 KW Isomerase; Complete proteome.
 SQ SEQUENCE 128 AA; 14793 MW; D14491B1283783E CRC64;

Query Match 0.9%; Score 7; DB 16; Length 128;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 660 DUFATIK 666
 |||||
 Db 84 DUFATIK 90

RESULT 90

Q9YAF7 PRELIMINARY; PRT; 128 AA.
 AC Q9YAF7;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Hypothetical protein APE1982.
 GN APE1982.
 OS Aeropyrum pernix.
 OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcaceae;
 OC Desulfurococcaceae; Aeropyrum.
 OX NCBI_TaxID=56636;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K1;
 RC MEDLINE=99310339; PubMed=10382966;
 RX Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
 RA Jinno K., Takahashi M., Sekine M., Baba S.-I., Akai A., Kosugi H.,
 RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
 RA Yamazaki J., Kushiida N., Oguchi A., Aoki K.-I., Kubota K.,
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.,
 RT "Complete genome sequence of an aerobic hyper-thermophilic
 RT crenarchaeon, Aeropyrum pernix K1."
 RL DNA Res. 6:83-101(1999).
 DR EMBL; AP000063; BAA80992.1;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 128 AA; 13557 MW; 9248C96789921467 CRC64;

Query Match 0.9%; Score 7; DB 17; Length 128;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 785 SNPSVS 791
 |||||
 Db 43 SNPSVS 49

RESULT 91

Q9XBD3 PRELIMINARY; PRT; 129 AA.
 AC Q9XBD3;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Mult-like protein.
 GN SC03260 OR SCE39.10.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycinae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RC Oliver K., Harris D.,
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RC Bentley S.D., Parkhill J., Barrett B.G., Rajandream N.A.,
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RC MEDLINE=97000351; PubMed=8843436;
 RX Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
 RA Kinash H., Hopwood D.A.,
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
 RL Mol. Microbiol. 21:77-96(1996).

[4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21966410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitch E., Rajandream M.A., Ruberford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL: AL939115; CAB40318.1; -.
DR HSSP: P08337; ITDM.
DR InterPro: IPR000086; NUDIX_hydrolase.
DR Pfam: PF00293; NUDIX; 1.
DR PRINTS: PR00502; NUDIXFAMILY.
DR PROSITE: PS00893; NUDIX; 1.
KW Complete proteome.
SQ SEQUENCE 129 AA; 14108 MW; 1BD428B6F66D1911 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 16; Length 129;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 LTPDEVS 39
DB 89 LTPDEVS 95
P74726 PRELIMINARY; PRT; 130 AA.
P74726
AC P74726;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein slr0588.
DE SLR0588.
OS Synecocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugita M., Saeemoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naito K., Okumura S.,
RA Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synecocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL: D90917; BAA18846.1; -.
SQ SEQUENCE 130 AA; 14238 MW; 5C35184E0301910 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 16; Length 130;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 268 DSLKQL 274
DB 111 DSLKQL 117

ID 069802 PRELIMINARY; PRT; 131 AA.
AC 069802;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Orl31.
GN Orl31.
OS Streptomyces ambifaciens.
OG Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomyces; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1889;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15154;
RX MEDLINE=89265212; PubMed=2657820;
RA Boccard F., Smokvina T., Pernodet J.L., Friedmann A., Guerinneau M.;
RT "Structural analysis of loci involved in pSAM2 site-specific
RT integration in Streptomyces.";
RL Plasmid 21:59-70(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15154;
RX MEDLINE=9374848; PubMed=8366038;
RA Hagege J., Pernodet J.L., Friedmann A., Guerinneau M.;
RT "Mode and origin of replication of pSAM2, a conjugative integrating
RT element of Streptomyces ambifaciens.";
RL Mol. Microbiol. 10:799-812(1993).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15154;
RX MEDLINE=95020551; PubMed=7934842;
RA Hagege J., Pernodet J.L., Gerbaud C., Sezonov G., Friedmann A.,
RA Guerinneau M.;
RT "Transfer function of the conjugative integrating element pSAM2 from
RT Streptomyces ambifaciens: characterization of a k11-kor system
RT associated with transfer.";
RL J. Bacteriol. 173:5529-5538(1993).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15154;
RX MEDLINE=94302137; PubMed=8029324;
RA Hagege J., Boccard F., Smokvina T., Pernodet J.L., Friedmann A.,
RA Guerinneau M.;
RT "Identification of a gene encoding the replication initiator protein
RT of the Streptomyces integrating element, pSAM2.";
RL Plasmid 31:166-183(1994).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15154;
RX MEDLINE=96100452; PubMed=8559072;
RA Sezonov G., Hagege J., Pernodet J.L., Friedmann A., Guerinneau M.;
RT "Characterization of prx, a gene for replication control in pSAM2, the
RT integrating element of Streptomyces ambifaciens.";
RL Mol. Microbiol. 17:533-544(1995).
RN [7]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15154;
RX MEDLINE=98292736; PubMed=9620953;
RA Sezonov G., Duchene A.M., Friedmann A., Guerinneau M., Pernodet J.L.;
RT "Replicase, excisionase, and integrase genes of the Streptomyces
RT element pSAM2 constitute an operon positively regulated by the prx
RT gene.";
RL J. Bacteriol. 180:3056-3061(1998).
DR EMBL: AJ005260; CAA06447.1; -.

```

DR InterPro: IPR000086; NUDIX_hydrolase.
DR Pfam: PF00293; NUDIX.1.
DR PRINTS: PR00502; NUDIXFAMILY.
DR PROSITE: PS00893; NUDIX.1.
KM plasmid.
SQ SEQUENCE 131 AA; 14295 MW; E44DC7C113475BD1 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 33 LTPDEVS 39
    |||||
Db 91 LTPDEVS 97

RESULT 94
O92P78 PRELIMINARY; PRT; 131 AA.
ID O92P78;
AC O92P78;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE Putative zinc uptake regulation protein.
GN ZUR OR R01912 OR SMC04242.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium.
OK NCBI_TaxId=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gonzy J., Bothe G., Ampe F., Barut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreao S., Gloux S.,
RA Godie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masny D.,
RA Pohl T., Portetel D., Puhler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL: AL591788; CAC46491.1; -.
DR InterPro: IPR002481; FUR.
DR Pfam: PF01475; FUR.1.
DR Prodom: PD002003; FUR.1.
KM Complete proteome.
SQ SEQUENCE 131 AA; 14643 MW; 8C8CCDA1B3E4CA88 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 466 ALDKLLE 472
    |||||
Db 47 ALDKLLE 53

RESULT 95
O86762 PRELIMINARY; PRT; 131 AA.
ID O86762;
AC O86762;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Mult-1 like protein.
GN SC05637 OR SC6A9.30C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomycetes.
OK NCBI_TaxId=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
```

```

RA Murphy L., Harris D.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redendach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmid and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2)."
RL Nature 417:141-147(2002).
DR EMBL: AL939124; CAA19915.1; -.
DR InterPro: IPR000086; NUDIX_hydrolase.
DR Pfam: PF00293; NUDIX.1.
DR PRINTS: PR00502; NUDIXFAMILY.
DR PROSITE: PS00893; NUDIX.1.
KM Complete proteome.
SQ SEQUENCE 131 AA; 14325 MW; F9F9BCB7C82AE791 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 33 LTPDEVS 39
    |||||
Db 91 LTPDEVS 97

RESULT 96
O9YL97 PRELIMINARY; PRT; 134 AA.
ID O9YL97;
AC O9YL97;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE PIX protein.
OS Human adenovirus type 9.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OK NCBI_TaxId=10527;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99174004; PubMed=10074157;
RA Thomas D.L., Shin S., Jiang B.H., Vogel H., Ross M.A., Kapilt M.,
RA Shenk T.E., Javlier R.T.;
RT "Early region 1 transforming functions are dispensable for mammary
RT tumorigenesis by human adenovirus type 9."
RL J. Virol. 73:3071-3079(1999).
DR EMBL: AF099665; AAD16306.1; -.
DR InterPro: IPR005641; Adeno_PIX.
DR Pfam: PF03955; Adeno_PIX.1.
SQ SEQUENCE 134 AA; 13751 MW; 39EB34E85C6CAD25 CRC64;
```

Query Match 0.9%; Score 7; DB 12; Length 134;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 776 EKLALL 782
DB 99 EKLALL 105

RESULT 97

QY 090640 PRELIMINARY; PRT; 134 AA.

AC 090640;
DT 01-NOV-1998 (TEMBLrel. 08, Created)
DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE Envelope glycoprotein (Fragment).
GN ENV.
OS Simian-Human immunodeficiency virus.
OC Viruses; Retrovird viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=57667;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SHIV-C2/1;
RA MEDLINE=99281901; PubMed=10355770;
RA Shinohara K., Sakai K., Ando S., Yoshino N., Takahashi E.,
RA Someya K., Suzuki Y., Nakasone T., Sasaki Y., Kaizu M., Lu Y.,
RA Honda M.;
RT "A highly pathogenic simian/human immunodeficiency virus with genetic
changes in cynomolgus monkey."
RL J. Gen. Virol. 80:1231-1240(1999).
DR EMBL; AF082832; AAC34101.1; -;
DR InterPro; IPR000328; Env_Gp41.
DR Pfam; PF00517; Gp41; 2. Env_Gp41.
KW Transmembrane.
FT NON_TER
SQ SEQUENCE 134 AA; 15700 MW; CB92BE2FDB6C094A CRC64;

Query Match 0.9%; Score 7; DB 15; Length 134;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 750 ATETLG 756
DB 96 ATETLG 102

RESULT 98

QY 08PN57 PRELIMINARY; PRT; 134 AA.

AC 08PN57;
DT 01-OCT-2002 (TEMBLrel. 22, Created)
DT 01-OCT-2002 (TEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TEMBLrel. 22, Last annotation update)
DE Hypothetical protein XAC1216.
GN XAC1216.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RA MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferris J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Canavaral F., Cardozo J., Chambergo F., Chaplin L.P.,
RA Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Petro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,

RA Martins E.C., Melanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira B.C., Tezza R.I.D.,
RA Trindade dos Santos M., Trufi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
host specificities."
RL Nature 417:459-463(2002).
DR EMBL; AF011751; AAM36088.1; -;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 134 AA; 14831 MW; 87247E2B424ECF8F CRC64;

Query Match 0.9%; Score 7; DB 16; Length 134;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 LKDAHA 123
DB 101 LKDAHA 107

RESULT 99

QY 08VS19 PRELIMINARY; PRT; 136 AA.

AC 08VS19;
DT 01-MAR-2002 (TEMBLrel. 20, Created)
DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Hypothetical 15.2 kDa protein.
GN CP0106.
OS Shigella flexneri 2a.
OC Plasmid pcP301.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=42897;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=301;
RA Jin Q., Zhang J.Y., Liu H., Yang J., Yang F., Zhang X.B., Wang J.H.,
RA Yang G.W., Wu H.T., Dong J., Sun L.L., Xue Y., Zhao A.L., Gao Y.S.,
RA Zhu J.P., Kan B., Chen S.X., Yao Z.J., He B.K., Chen R.S., Ma D.L.,
RA Yuan Z.H., Xu J.G., Wang Y., Shen Y., Lu W.C., Qiang B.Q., Wen Y.M.,
RA Hou Y.D.;
RT "Complete DNA sequence and analysis of the large virulence plasmid
PCP301 of Shigella flexneri."
RL Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF386526; AAL72401.1; -;
DR InterPro; IPR002514; Transposase_8.
DR Pfam; PF01527; Transposase_8; 1.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 136 AA; 15159 MW; 8E2A0FA917B3DF CRC64;

Query Match 0.9%; Score 7; DB 2; Length 136;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 202 ASELAA 208
DB 85 ASELAA 91

RESULT 100

QY 017947 PRELIMINARY; PRT; 136 AA.

AC 017947;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE C13G3.1 protein.
GN C13G3.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;

OC Rhabditiidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA McMurray A.A.;
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 investigating biology.";
 RL Science 282:2012-2018(1998).
 DR EMBL; 274027; CAA98421.1; -.
 DR WormPep; C13G3.1; CE05277.
 SQ SEQUENCE 136 AA; 14914 MW; 804BDBAB75A6645F CRC64;
 Query Match 0.9%; Score 7; DB 5; Length 136;
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 783 KGSNPSS 789
 |||||
 Db 96 KGSNPSS 102

Search completed: November 14, 2003, 11:01:14
 Job time : 109 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 10:58:50 ; Search time 112 Seconds
(without alignments)
300.709 Million cell updates/sec

Title: US-09-765-271-56

Perfect score: 796
Sequence: 1 SYELGLYQARTVKNRVS.....KLIALIKGNPSVSKKIN 796

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 200 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/backfill1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	795	99.9	796	3	US-08-961-083-56
2	795	99.9	796	4	US-09-536-784-56
3	695	87.3	819	4	US-09-468-656A-8
4	60	7.5	819	4	US-09-468-656A-10
5	60	7.5	838	4	US-09-468-656A-4
6	51	6.4	763	3	US-08-961-083-66
7	51	6.4	763	4	US-09-536-784-66
8	32	4.0	447	3	US-08-961-083-182
9	32	4.0	447	4	US-09-536-784-182
10	32	4.0	484	4	US-09-468-656A-6
11	9	1.1	795	4	US-09-328-352-6143
12	9	1.1	921	1	US-08-188-582-2
13	9	1.1	921	1	US-08-946-715-2
14	8	1.0	110	2	US-08-569-166-34
15	8	1.0	332	4	US-09-107-532A-5477
16	8	1.0	490	4	US-09-134-001C-5116
17	8	1.0	503	3	US-08-999-774A-10
18	7	0.9	15	1	US-08-221-583-57
19	7	0.9	15	1	US-08-221-583-58
20	7	0.9	15	1	US-08-221-583-59
21	7	0.9	15	5	PCT-US95-04018-57
22	7	0.9	15	5	PCT-US95-04018-58
23	7	0.9	15	5	PCT-US95-04018-59
24	7	0.9	27	3	US-09-020-116-4
25	7	0.9	27	3	US-09-608-902-4
26	7	0.9	45	3	US-09-020-116-2
27	7	0.9	45	4	US-09-608-902-2

28	7	0.9	50	3	US-09-171-646-1	Sequence 1, Appl1
29	7	0.9	66	3	US-09-020-116-1	Sequence 1, Appl1
30	7	0.9	66	4	US-09-608-902-1	Sequence 1, Appl1
31	7	0.9	76	4	US-09-328-352-6019	Sequence 6019, Ap
32	7	0.9	117	4	US-09-134-001C-4585	Sequence 4585, Ap
33	7	0.9	149	4	US-09-252-991A-28698	Sequence 28698, A
34	7	0.9	158	2	US-08-917-456-2	Sequence 2, Appl1
35	7	0.9	158	3	US-09-229-804-2	Sequence 2, Appl1
36	7	0.9	163	4	US-09-252-991A-30622	Sequence 20622, A
37	7	0.9	191	4	US-09-252-991A-30217	Sequence 30217, A
38	7	0.9	198	4	US-09-613-303-35	Sequence 35, Appl
39	7	0.9	201	2	US-08-916-901-3	Sequence 3, Appl1
40	7	0.9	201	2	US-08-916-901-8	Sequence 8, Appl1
41	7	0.9	201	4	US-09-154-602-3	Sequence 3, Appl1
42	7	0.9	201	4	US-09-154-602-8	Sequence 8, Appl1
43	7	0.9	226	3	US-08-908-332-5	Sequence 5, Appl1
44	7	0.9	235	4	US-09-326-394-4	Sequence 4, Appl1
45	7	0.9	235	4	US-09-580-235-2	Sequence 2, Appl1
46	7	0.9	235	4	US-09-580-235-6	Sequence 6, Appl1
47	7	0.9	235	4	US-09-580-235-8	Sequence 8, Appl1
48	7	0.9	235	4	US-09-580-181-2	Sequence 2, Appl1
49	7	0.9	235	4	US-09-580-181-4	Sequence 4, Appl1
50	7	0.9	235	4	US-09-580-181-6	Sequence 6, Appl1
51	7	0.9	235	4	US-09-580-181-8	Sequence 8, Appl1
52	7	0.9	235	4	US-09-102-530-2	Sequence 2, Appl1
53	7	0.9	235	4	US-09-102-530-4	Sequence 4, Appl1
54	7	0.9	235	4	US-09-102-530-6	Sequence 6, Appl1
55	7	0.9	235	4	US-09-102-530-8	Sequence 8, Appl1
56	7	0.9	235	4	US-09-620-405B-487	Sequence 487, App
57	7	0.9	235	4	US-09-620-405B-487	Sequence 487, App
58	7	0.9	235	4	US-09-604-287A-487	Sequence 487, App
59	7	0.9	240	4	US-09-107-533A-6634	Sequence 6634, Ap
60	7	0.9	245	4	US-09-252-991A-29404	Sequence 29404, A
61	7	0.9	257	4	US-09-579-845-10	Sequence 10, Appl
62	7	0.9	261	1	US-08-622-353-2	Sequence 2, Appl1
63	7	0.9	261	2	US-08-622-353A-2	Sequence 2, Appl1
64	7	0.9	261	3	US-08-826-390-2	Sequence 2, Appl1
65	7	0.9	298	4	US-09-252-991A-17944	Sequence 17944, A
66	7	0.9	305	4	US-09-252-991A-31029	Sequence 31029, A
67	7	0.9	312	4	US-09-252-991A-29807	Sequence 29807, A
68	7	0.9	325	4	US-09-198-452A-745	Sequence 745, App
69	7	0.9	383	4	US-09-459-749D-17	Sequence 17, Appl
70	7	0.9	390	3	US-08-861-774E-82	Sequence 82, Appl
71	7	0.9	406	4	US-08-311-731A-12	Sequence 12, Appl
72	7	0.9	447	4	US-09-252-991A-22113	Sequence 22113, A
73	7	0.9	461	1	US-08-385-229-2	Sequence 2, Appl1
74	7	0.9	461	2	US-08-650-000-2	Sequence 2, Appl1
75	7	0.9	461	3	US-09-042-785A-7	Sequence 7, Appl1
76	7	0.9	461	3	US-08-477-347-3	Sequence 3, Appl1
77	7	0.9	461	3	US-09-006-353A-4	Sequence 4, Appl1
78	7	0.9	461	3	US-08-476-862-2	Sequence 2, Appl1
79	7	0.9	461	4	US-09-573-986-4	Sequence 4, Appl1
80	7	0.9	461	4	US-08-406-824A-2	Sequence 2, Appl1
81	7	0.9	461	4	US-09-800-909-2	Sequence 2, Appl1
82	7	0.9	461	4	US-09-758-124-2	Sequence 2, Appl1
83	7	0.9	461	6	5395760-2	Patent No. 5395760
84	7	0.9	465	4	US-09-328-352-4742	Sequence 4742, Ap
85	7	0.9	486	4	US-09-252-991A-17797	Sequence 17797, A
86	7	0.9	486	1	US-08-243-010-1	Sequence 1, Appl1
87	7	0.9	498	4	US-09-252-991A-27183	Sequence 27183, A
88	7	0.9	505	4	US-09-620-405B-478	Sequence 478, App
89	7	0.9	505	4	US-09-604-287A-478	Sequence 478, App
90	7	0.9	505	4	US-09-604-287A-485	Sequence 485, App
91	7	0.9	506	4	US-09-252-991A-19121	Sequence 19121, A
92	7	0.9	511	3	US-08-991-677-4	Sequence 4, Appl1
93	7	0.9	511	3	US-08-426-509A-16	Sequence 16, Appl
94	7	0.9	512	4	US-08-232-545-16	Sequence 16, Appl
95	7	0.9	512	4	PCT-US95-05008-16	Sequence 16, Appl
96	7	0.9	513	5	US-09-564-808-6	Sequence 6, Appl1
97	7	0.9	513	4	US-08-385-229-4	Sequence 4, Appl1
98	7	0.9	518	1	US-09-579-845-1	Sequence 1, Appl1
99	7	0.9	518	4		
100	7	0.9				

101	7	0.9	518	4	US-09-579-845-3	Sequence 3, Appli
102	7	0.9	520	2	US-09-091-433-2	Sequence 2, Appli
103	7	0.9	520	4	US-09-387-663-2	Sequence 2, Appli
104	7	0.9	520	4	US-09-564-808-2	Sequence 2, Appli
105	7	0.9	520	4	US-09-564-808-4	Sequence 4, Appli
106	7	0.9	528	4	US-09-252-991A-1812	Sequence 18132, A
107	7	0.9	529	4	US-09-615-192A-405	Sequence 405, App
108	7	0.9	530	4	US-09-252-991A-21963	Sequence 21963, A
109	7	0.9	539	1	US-08-399-666-122	Sequence 122, App
110	7	0.9	588	1	US-08-460-860-4	Sequence 4, Appli
111	7	0.9	596	4	US-09-252-991A-18934	Sequence 18934, A
112	7	0.9	597	1	US-08-399-666-102	Sequence 102, App
113	7	0.9	649	3	US-09-295-186-18	Sequence 18, Appl
114	7	0.9	680	4	US-09-252-991A-17566	Sequence 17566, A
115	7	0.9	688	4	US-09-252-991A-17969	Sequence 17969, A
116	7	0.9	717	4	US-09-107-532A-6277	Sequence 6277, Ap
117	7	0.9	801	4	US-09-252-991A-22729	Sequence 22729, A
118	7	0.9	832	4	US-09-252-991A-22753	Sequence 22753, A
119	7	0.9	864	4	US-09-323-872A-28	Sequence 28, Appl
120	7	0.9	864	4	US-09-072-433-16	Sequence 16, Appl
121	7	0.9	1193	1	US-08-317-450B-13	Sequence 13, Appl
122	7	0.9	1193	3	US-08-800-593-13	Sequence 13, Appl
123	7	0.9	1248	2	US-08-348-353-17	Sequence 17, Appl
124	7	0.9	1248	2	US-08-465-965-17	Sequence 17, Appl
125	7	0.9	1248	3	US-08-465-966-17	Sequence 17, Appl
126	7	0.9	1288	4	US-09-546-934-4	Sequence 4, Appli
127	7	0.9	1289	4	US-09-546-934-1	Sequence 1, Appli
128	7	0.9	2547	3	US-09-058-489-35	Sequence 35, Appl
129	7	0.9	2736	4	US-09-252-991A-30227	Sequence 30227, A
130	6	0.8	7	4	US-09-187-859-3833	Sequence 3833, Ap
131	6	0.8	7	4	US-09-839-542B-3833	Sequence 3833, Ap
132	6	0.8	8	4	US-09-187-859-3834	Sequence 3834, Ap
133	6	0.8	8	4	US-09-839-542B-3834	Sequence 15, Appl
134	6	0.8	10	3	US-08-884-324-15	Sequence 55, Appl
135	6	0.8	10	3	US-09-106-216-55	Sequence 7, Appli
136	6	0.8	10	3	US-08-832-180-7	Sequence 1, Appli
137	6	0.8	10	3	US-08-832-198-1	Sequence 1, Appli
138	6	0.8	10	4	US-09-819-902-1	Sequence 1, Appli
139	6	0.8	10	4	US-09-752-510-1	Sequence 27, Appl
140	6	0.8	10	4	US-08-982-285-27	Sequence 3, Appli
141	6	0.8	11	3	US-08-558-818-3	Sequence 3, Appli
142	6	0.8	11	3	US-08-974-469A-3	Sequence 3, Appli
143	6	0.8	11	4	US-09-025-596-44	Sequence 44, Appl
144	6	0.8	11	4	US-09-711-899-3	Sequence 3, Appli
145	6	0.8	12	2	US-08-739-015-1	Sequence 1, Appli
146	6	0.8	12	2	US-08-741-663-1	Sequence 1, Appli
147	6	0.8	13	3	US-08-832-198-7	Sequence 7, Appli
148	6	0.8	13	4	US-09-819-902-7	Sequence 7, Appli
149	6	0.8	13	4	US-09-752-510-7	Sequence 44, Appl
150	6	0.8	15	3	US-08-448-194-44	Sequence 21, Appl
151	6	0.8	15	4	US-08-867-921-44	Sequence 21, Appl
152	6	0.8	17	4	US-08-472-701-21	Sequence 124, App
153	6	0.8	17	4	US-09-690-454-124	Sequence 20, Appl
154	6	0.8	17	5	PCT-US95-08596-21	Sequence 13, Appl
155	6	0.8	17	5	PCT-US95-08621-20	Sequence 12, Appl
156	6	0.8	19	3	US-09-171-878-13	Sequence 40, Appl
157	6	0.8	19	4	US-09-307-143-12	Sequence 3, Appli
158	6	0.8	21	4	US-09-387-418A-40	Sequence 1147, Ap
159	6	0.8	29	3	US-09-023-731-3	Sequence 58, Appl
160	6	0.8	29	4	US-09-205-258-1147	Sequence 58, Appl
161	6	0.8	31	1	US-08-614-935-58	Sequence 59, Appl
162	6	0.8	31	1	US-08-614-935-59	Sequence 61, Appl
163	6	0.8	31	3	US-09-130-287-58	Sequence 66, Appl
164	6	0.8	31	3	US-09-130-287-59	Sequence 59, Appl
165	6	0.8	32	4	US-09-053-197A-61	Sequence 59, Appl
166	6	0.8	32	4	US-09-085-761A-66	Sequence 66, Appl
167	6	0.8	35	1	US-08-038-343A-5	Sequence 1, Appli
168	6	0.8	35	1	US-08-366-690-1	Sequence 1, Appli
169	6	0.8	36	4	US-09-227-357-182	Sequence 182, App
170	6	0.8	40	1	US-08-313-075A-58	Sequence 58, Appl
171	6	0.8	41	4	US-07-145-002B-65	Sequence 21, Appl
172	6	0.8	49	3	US-09-004-406C-21	Sequence 2, Appli
173	6	0.8	50	3	US-08-832-198-2	

174	6	0.8	50	4	US-09-819-902-2	Sequence 2, Appli
175	6	0.8	50	4	US-09-752-510-2	Sequence 2, Appli
176	6	0.8	53	4	US-08-861-476C-8	Sequence 8, Appli
177	6	0.8	57	3	US-09-023-731-4	Sequence 4, Appli
178	6	0.8	60	1	US-08-370-225-32	Sequence 32, Appl
179	6	0.8	60	1	US-08-461-859-32	Sequence 32, Appl
180	6	0.8	60	3	US-08-461-607-11	Sequence 11, Appl
181	6	0.8	60	3	US-09-363-600-11	Sequence 11, Appl
182	6	0.8	60	5	PCT-US93-10069-32	Sequence 32, Appl
183	6	0.8	63	1	US-08-447-411-23	Sequence 61, Appl
184	6	0.8	63	1	US-08-447-411-61	Sequence 62, Appl
185	6	0.8	63	2	US-08-447-411-62	Sequence 18, Appl
186	6	0.8	63	2	US-08-662-227-18	Sequence 18, Appl
187	6	0.8	63	2	US-08-662-227-19	Sequence 19, Appl
188	6	0.8	63	4	US-09-017-947-18	Sequence 19, Appl
189	6	0.8	63	4	US-09-017-947-19	Sequence 19, Appl
190	6	0.8	64	4	US-09-722-210-491	Sequence 491, App
191	6	0.8	69	2	US-08-245-511-27	Sequence 27, Appl
192	6	0.8	69	2	US-08-600-993A-27	Sequence 27, Appl
193	6	0.8	71	1	US-08-261-677-13	Sequence 13, Appl
194	6	0.8	71	1	US-08-384-556A-9	Sequence 9, Appli
195	6	0.8	71	2	US-08-331-355A-13	Sequence 13, Appl
196	6	0.8	71	5	PCT-US94-12364-13	Sequence 13, Appl
197	6	0.8	71	5	PCT-US95-07753-9	Sequence 9, Appli
198	6	0.8	74	3	US-08-938-669A-31	Sequence 31, Appl
199	6	0.8	74	4	US-09-306-828-11	Sequence 31, Appl
200	6	0.8	78	4	US-09-107-532A-4405	Sequence 4405, Ap

RESULT 1

US-08-961-083-56

Sequence 56, Application US/08961083

Patent No. 6159469

GENERAL INFORMATION:

APPLICANT: Chol et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines

NUMBER OF SEQUENCES: 452

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/961,083

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8512

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 56:

SEQUENCE CHARACTERISTICS:

LENGTH: 796 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-961-083-56

Query Match 99.9%; Score 795; DB 3; Length 796;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 SYELGIYQARTYKNNRVSYIDGKQATQKTENLTPEVSKREGINAEOIVIKITDGGYVT 60
DB 1 SYELGIYQARTYKNNRVSYIDGKQATQKTENLTPEVSKREGINAEOIVIKITDGGYVT 60
QY 61 SHGDHYHYNGKVPYDAIISSEILMKDPYKDKEDIVNEVGGYIKVDGKYVYLKDA 120
DB 61 SHGDHYHYNGKVPYDAIISSEILMKDPYKDKEDIVNEVGGYIKVDGKYVYLKDA 120
QY 121 AHADVRTKEEINRQKQEHSGHREGTPRNDGAVALARSGRYTTDDGYIFNADIIEDT 180
DB 121 AHADVRTKEEINRQKQEHSGHREGTPRNDGAVALARSGRYTTDDGYIFNADIIEDT 180
QY 181 GDAYIVPHGDHYHYPKNELSASELAAPLISGRGNLSNRTYRRQNSDNTSRITWVPS 240
DB 181 GDAYIVPHGDHYHYPKNELSASELAAPLISGRGNLSNRTYRRQNSDNTSRITWVPS 240
QY 241 VSNPGTTNTNTSNNTNSQASQNSDIDSLKQLYKLPSQRHVESDGLVPFPAQITTSRT 300
DB 241 VSNPGTTNTNTSNNTNSQASQNSDIDSLKQLYKLPSQRHVESDGLVPFPAQITTSRT 300
QY 301 ARGVAVPHGDHYHFIPIYSQMSSELEERIAIIPLRYSNHWVPDSRPEQSPQPTPEPSPG 360
DB 301 ARGVAVPHGDHYHFIPIYSQMSSELEERIAIIPLRYSNHWVPDSRPEQSPQPTPEPSPG 360
QY 361 POPAPNLKIDSNSLSVSQLVRKVGEGYVEEKGISRYVFAKOLPSETVNLNLSKLSKQBS 420
DB 361 POPAPNLKIDSNSLSVSQLVRKVGEGYVEEKGISRYVFAKOLPSETVNLNLSKLSKQBS 420
QY 421 VSHITLAKKENAVAPRDQEFYDKAYNLITFAHAKALFYNKGKNSDFQALDKLRLNDESTN 480
DB 421 VSHITLAKKENAVAPRDQEFYDKAYNLITFAHAKALFYNKGKNSDFQALDKLRLNDESTN 480
QY 481 KEKLVDDLLAFAPITHPERLGRKPNQSIETDEDEVIAQADKYTTSDDGYIFDEHDIISD 540
DB 481 KEKLVDDLLAFAPITHPERLGRKPNQSIETDEDEVIAQADKYTTSDDGYIFDEHDIISD 540
QY 541 EGDAYVTPHMGSHWIGKDSLSDKEKVAAQAYTKEKGLPPSPBDADVKNPFGDSAAIY 600
DB 541 EGDAYVTPHMGSHWIGKDSLSDKEKVAAQAYTKEKGLPPSPBDADVKNPFGDSAAIY 600
QY 601 NRKVGKRIPLVRLPYMVEHTEVEKNGNLIIPKDHYNIKFAMFDDHTYKAPNGYTTLED 660
DB 601 NRKVGKRIPLVRLPYMVEHTEVEKNGNLIIPKDHYNIKFAMFDDHTYKAPNGYTTLED 660
QY 661 LFAATIKYVYEHDERPHSNDGNGNASEHILGKKHSDENKNFKADEEVEETPAPEVP 720
DB 661 LFAATIKYVYEHDERPHSNDGNGNASEHILGKKHSDENKNFKADEEVEETPAPEVP 720
QY 721 QVETEKVEAQLKEAEVLLAKVTDSILKANATETLAGLRNNLLQIMDNNSIMAEAKLIA 780
DB 721 QVETEKVEAQLKEAEVLLAKVTDSILKANATETLAGLRNNLLQIMDNNSIMAEAKLIA 780
QY 781 LKGSNPSSVSXEKIN 796
DB 781 LKGSNPSSVSXEKIN 796
```

RESULT 2
US-09-536-784-56

; Sequence 56, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESS: Human Genome Sciences, Inc.

```
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/536,784
FILING DATE: 30-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: OCT-30-1997
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB340P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 796 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 56:
```

US-09-536-784-56

Query Match 99.9%; Score 795; DB 4; Length 796;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 SYELGIYQARTYKNNRVSYIDGKQATQKTENLTPEVSKREGINAEOIVIKITDGGYVT 60
DB 1 SYELGIYQARTYKNNRVSYIDGKQATQKTENLTPEVSKREGINAEOIVIKITDGGYVT 60
QY 61 SHGDHYHYNGKVPYDAIISSEILMKDPYKDKEDIVNEVGGYIKVDGKYVYLKDA 120
DB 61 SHGDHYHYNGKVPYDAIISSEILMKDPYKDKEDIVNEVGGYIKVDGKYVYLKDA 120
QY 121 AHADVRTKEEINRQKQEHSGHREGTPRNDGAVALARSGRYTTDDGYIFNADIIEDT 180
DB 121 AHADVRTKEEINRQKQEHSGHREGTPRNDGAVALARSGRYTTDDGYIFNADIIEDT 180
QY 181 GDAYIVPHGDHYHYPKNELSASELAAPLISGRGNLSNRTYRRQNSDNTSRITWVPS 240
DB 181 GDAYIVPHGDHYHYPKNELSASELAAPLISGRGNLSNRTYRRQNSDNTSRITWVPS 240
QY 241 VSNPGTTNTNTSNNTNSQASQNSDIDSLKQLYKLPSQRHVESDGLVPFPAQITTSRT 300
DB 241 VSNPGTTNTNTSNNTNSQASQNSDIDSLKQLYKLPSQRHVESDGLVPFPAQITTSRT 300
QY 301 ARGVAVPHGDHYHFIPIYSQMSSELEERIAIIPLRYSNHWVPDSRPEQSPQPTPEPSPG 360
DB 301 ARGVAVPHGDHYHFIPIYSQMSSELEERIAIIPLRYSNHWVPDSRPEQSPQPTPEPSPG 360
QY 361 POPAPNLKIDSNSLSVSQLVRKVGEGYVEEKGISRYVFAKOLPSETVNLNLSKLSKQBS 420
DB 361 POPAPNLKIDSNSLSVSQLVRKVGEGYVEEKGISRYVFAKOLPSETVNLNLSKLSKQBS 420
QY 421 VSHITLAKKENAVAPRDQEFYDKAYNLITFAHAKALFYNKGKNSDFQALDKLRLNDESTN 480
DB 421 VSHITLAKKENAVAPRDQEFYDKAYNLITFAHAKALFYNKGKNSDFQALDKLRLNDESTN 480
QY 481 KEKLVDDLLAFAPITHPERLGRKPNQSIETDEDEVIAQADKYTTSDDGYIFDEHDIISD 540
DB 481 KEKLVDDLLAFAPITHPERLGRKPNQSIETDEDEVIAQADKYTTSDDGYIFDEHDIISD 540
```

```

Qy 541 EGDAYVTPHMGSHWIGKDSLSDKEKVAQAAYTKEKGLPPSPDADVANKPTGDSAAIY 600
Db 541 EGDAYVTPHMGSHWIGKDSLSDKEKVAQAAYTKEKGLPPSPDADVANKPTGDSAAIY 600
Qy 601 NRKVGKRIPLVRLPYWVEHTVEYKGNLIIPHKDHYNIKFAWFDHTTYKAPNGYTLSD 660
Db 601 NRKVGKRIPLVRLPYWVEHTVEYKGNLIIPHKDHYNIKFAWFDHTTYKAPNGYTLSD 660
Qy 661 LFATIKYVVEHPDERPHSDNGMNASSEHVLGKDHSEDPNKNFKADEEPVEETPAEPVP 720
Db 661 LFATIKYVVEHPDERPHSDNGMNASSEHVLGKDHSEDPNKNFKADEEPVEETPAEPVP 720
Qy 721 QVETEKVAQLKEAEVLLAKYTDSSLKANATETLAGLRNNLTLOIMDNNSIMAEKLLA 780
Db 721 QVETEKVAQLKEAEVLLAKYTDSSLKANATETLAGLRNNLTLOIMDNNSIMAEKLLA 780
Qy 781 LKGSNPSSVSKEKIN 796
Db 781 LKGSNPSSVSKEKIN 796

```

```

RESULT 3
US-09-468-656A-8
; Sequence 8, Application US/09468656A
; Patent No. 6582706
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
; FILE REFERENCE: 469201-444
; CURRENT APPLICATION NUMBER: US/09/468,656A
; CURRENT FILING DATE: 1998-12-02
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 819
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-468-656A-8

```

```

Query Match 87.3%; Score 695; DB 4; Length 819;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 1 SYELGLVQARTVKNRNVSYIDGKQATKTEMLTPDEVSKREGINAEOIVIKITDQGYVT 60
Db 21 SYELGLVQARTVKNRNVSYIDGKQATKTEMLTPDEVSKREGINAEOIVIKITDQGYVT 80
Qy 61 SHGDHYHYNKGVPYDAIISELLMKDPNYKLKDEIDINEVKGVIKVDGKYVYLLKDA 120
Db 81 SHGDHYHYNKGVPYDAIISELLMKDPNYKLKDEIDINEVKGVIKVDGKYVYLLKDA 140
Qy 121 AAHADNRKKEIRINROKQSHSQHREGCTPRNDGAVLAASOGRYTTDDGYIFNASTIIDT 180
Db 141 AAHADNRKKEIRINROKQSHSQHREGCTPRNDGAVLAASOGRYTTDDGYIFNASTIIDT 200
Qy 181 GDAYIVPBGDHYHYIPKNELSASELAAABAFLSGNGNSNSTRYRONSDMSTRNWPS 240
Db 201 GDAYIVPBGDHYHYIPKNELSASELAAABAFLSGNGNSNSTRYRONSDMSTRNWPS 260
Qy 241 VSNPCTTNTSNNSTNSQASQSDNDIDSLKQLYKLPLSQHVESDGLVDPQAQITSRT 300
Db 261 VSNPCTTNTSNNSTNSQASQSDNDIDSLKQLYKLPLSQHVESDGLVDPQAQITSRT 320
Qy 301 ARGVAVPHGDHYHYIPYQMSLEIRIARIIPLRKRSNHWVPDSRPEQSPQPTPEBSPG 360
Db 321 ARGVAVPHGDHYHYIPYQMSLEIRIARIIPLRKRSNHWVPDSRPEQSPQPTPEBSPG 380

```

```

Qy 361 POPAPNLIKDSNSSVLSQVARKVGEVYFBEKGISRYVFAKDLBSETVKNLESKLSQES 420
Db 381 POPAPNLIKDSNSSVLSQVARKVGEVYFBEKGISRYVFAKDLBSETVKNLESKLSQES 440
Qy 421 VSHTLTAKKENVARROPEFDKAYNLLTEAHKALFANKGRSDPOALDKLERLNDESTN 480
Db 441 VSHTLTAKKENVARROPEFDKAYNLLTEAHKALFANKGRSDPOALDKLERLNDESTN 500
Qy 481 KEKLVDDLAFIAPITPERLGKPNISOIETDEVRJAOLADKTTSDGYIFDEHDIISD 540
Db 501 KEKLVDDLAFIAPITPERLGKPNISOIETDEVRJAOLADKTTSDGYIFDEHDIISD 560
Qy 541 EGDAYVTPHMGSHWIGKDSLSDKEKVAQAAYTKEKGLPPSPDADVANKPTGDSAAIY 600
Db 561 EGDAYVTPHMGSHWIGKDSLSDKEKVAQAAYTKEKGLPPSPDADVANKPTGDSAAIY 620
Qy 601 NRKVGKRIPLVRLPYWVEHTVEYKGNLIIPHKDHYNIKFAWFDHTTYKAPNGYTLSD 660
Db 621 NRKVGKRIPLVRLPYWVEHTVEYKGNLIIPHKDHYNIKFAWFDHTTYKAPNGYTLSD 680
Qy 661 LFATIKYVVEHPDERPHSDNGMNASSEHVLGKDHSEDPNKNFKADEEPVEETPAEPVP 720
Db 681 LFATIKYVVEHPDERPHSDNGMNASSEHVLGKDHSEDPNKNFKADEEPVEETPAEPVP 740
Qy 721 QVETEKVAQLKEAEVLLAKYTDSSLKANATETLAGLRNNLTLOIMDNNSIMAEKLLA 780
Db 741 QVETEKVAQLKEAEVLLAKYTDSSLKANATETLAGLRNNLTLOIMDNNSIMAEKLLA 800
Qy 781 LKGSNPSSVSKEKIN 796
Db 801 LKGSNPSSVSKEKIN 816

```

```

RESULT 4
US-09-468-656A-10
; Sequence 10, Application US/09468656A
; Patent No. 6582706
; GENERAL INFORMATION:
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
; FILE REFERENCE: 469201-444
; CURRENT APPLICATION NUMBER: US/09/468,656A
; CURRENT FILING DATE: 1998-12-02
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 819
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-468-656A-10

```

```

Query Match 7.5%; Score 60; DB 4; Length 819;
Best Local Similarity 100.0%; Pred. No. 1,3e-50;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 31 ENLTPDEVSKREGINAEOIVIKITDQGYVTSBGDHYHYNKGVPYDAIISELLMKDPNY 90
Db 52 ENLTPDEVSKREGINAEOIVIKITDQGYVTSBGDHYHYNKGVPYDAIISELLMKDPNY 111

```

```

RESULT 5
US-09-468-656A-4
; Sequence 4, Application US/09468656A
; Patent No. 6582706
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus

```

;; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
;; FILE OF INVENTION: Moflits
;; FILE REFERENCE: 469201-444
;; CURRENT APPLICATION NUMBER: US/09/468,656A
;; CURRENT FILING DATE: 1999-12-02
;; PRIOR APPLICATION NUMBER: 60/113,048
;; PRIOR FILING DATE: 1998-12-21
;; NUMBER OF SEQ ID NOS: 14
;; SOFTWARE: Patent In Ver. 2.1
;; SEQ ID NO: 4
;; LENGTH: 838
;; TYPE: PRT
;; ORGANISM: Streptococcus pneumoniae
US-09-468-656A-4

Query Match
Best Local Similarity 100.0%; Score 60; DB 4; Length 838;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 31 ENLTPDEVSKREGINAEQIVIKITDQGYTSHGHDHYHYKVPYDAISELLMDPNY 90
|||||
Db 52 ENLTPDEVSKREGINAEQIVIKITDQGYTSHGHDHYHYKVPYDAISELLMDPNY 111

RESULT 6
US-08-961-083-66
; Sequence 66, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 763 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-961-083-66

Query Match 6.4%; Score 51; DB 3; Length 763;
Best Local Similarity 100.0%; Pred. No. 1e-41;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 160 QGRYTTDDGYIFNASDIIEDTGDYIVPHGDHYHYIPKXELASSELAAEA 210
|||||

Db 159 QGRYTTDDGYIFNASDIIEDTGDYIVPHGDHYHYIPKXELASSELAAEA 209

RESULT 7
US-09-536-784-66
; Sequence 66, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 763 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 66:
US-09-536-784-66

Query Match 6.4%; Score 51; DB 4; Length 763;
Best Local Similarity 100.0%; Pred. No. 1e-41;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 160 QGRYTTDDGYIFNASDIIEDTGDYIVPHGDHYHYIPKXELASSELAAEA 210
|||||
Db 159 QGRYTTDDGYIFNASDIIEDTGDYIVPHGDHYHYIPKXELASSELAAEA 209

RESULT 8
US-08-961-083-182
; Sequence 182, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 182:
SEQUENCE CHARACTERISTICS:
LENGTH: 447 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-182

Query Match 4.0%; Score 32; DB 3; Length 447;
Best Local Similarity 100.0%; Pred. No. 3.9e-23;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 46 AEQIVIKITDGGVYTSQDHYHYNGKVPYDA 77
DB 43 AEQIVIKITDGGVYTSQDHYHYNGKVPYDA 74

RESULT 9
US-09-536-784-182
Sequence 182, Application US/09536784
Patent No. 6573082
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/536,784
FILING DATE: 30-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: OCT-30-1997
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB340P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 182:
SEQUENCE CHARACTERISTICS:
LENGTH: 447 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 182:
US-09-536-784-182

Query Match 4.0%; Score 32; DB 4; Length 447;
Best Local Similarity 100.0%; Pred. No. 3.9e-23;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 46 AEQIVIKITDGGVYTSQDHYHYNGKVPYDA 77
DB 43 AEQIVIKITDGGVYTSQDHYHYNGKVPYDA 74

RESULT 10
US-09-468-656A-6
Sequence 6, Application US/09468656A
Patent No. 6582706
GENERAL INFORMATION:
APPLICANT: Johnson, Leslie S.
TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
FILE REFERENCE: 469201-444
CURRENT APPLICATION NUMBER: US/09/468,656A
CURRENT FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: 60/113,048
PRIOR FILING DATE: 1998-12-21
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 6
LENGTH: 484
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-468-656A-6

Query Match 4.0%; Score 32; DB 4; Length 484;
Best Local Similarity 100.0%; Pred. No. 4.2e-23;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 46 AEQIVIKITDGGVYTSQDHYHYNGKVPYDA 77
DB 67 AEQIVIKITDGGVYTSQDHYHYNGKVPYDA 98

RESULT 11
US-09-328-352-6143
Sequence 6143, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Bretton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 6143
LENGTH: 795
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-6143

Query Match 1.1%; Score 9; DB 4; Length 795;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 201 SASFLAAE 209
DB 428 SASFLAAE 436

RESULT 12
US-08-188-582-2
; Sequence 2, Application US/08188582
; Patent No. 5534410
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/188,582
; FILING DATE: 28-JAN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 921 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-188-582-2

Query Match 1.1%; Score 9; DB 1; Length 921;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 TINTSNNSNT 257
Db 178 TINTSNNSNT 186

RESULT 13
US-08-646-715-2
; Sequence 2, Application US/08646715
; Patent No. 5637686
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE

NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/646,715
; FILING DATE: 09-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/188,582
; FILING DATE: 28-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 921 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-646-715-2

Query Match 1.1%; Score 9; DB 1; Length 921;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 TINTSNNSNT 257
Db 178 TINTSNNSNT 186

RESULT 14
US-08-569-166-34
; Sequence 34, Application US/08569166
; Patent No. 5830722
; GENERAL INFORMATION:
; APPLICANT: NICOLAS, LUC
; APPLICANT: CHARLES, JEAN-FRANCOIS
; APPLICANT: DELECLUSE, ARMELE
; APPLICANT: BARLOY, FREDERIQUE
; TITLE OF INVENTION: CLOSTRIDIUM BIFERMENTANS DNA FRAGMENT
; TITLE OF INVENTION: BEARING GENES CODING FOR PROTEINS LINKED TO AN
; TITLE OF INVENTION: INSECTICIDAL ACTIVITY
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/569,166

FILING DATE: 05-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR94/00768
FILING DATE: 24-JUN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 93/07795
FILING DATE: 25-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 660-106-0 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-3220
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 110 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-569-166-34

Query Match 1.0%; Score 8; DB 2; Length 110;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 250 NTSNNSNT 257
|||||
67 NTSNNSNT 74

Db

RESULT 15
US-09-107-532A-5477
Sequence 5477, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-8277
TELEFAX: (781)893-5007
INFORMATION FOR SEQ ID NO: 5477:
SEQUENCE CHARACTERISTICS:
LENGTH: 332 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...332
SEQUENCE DESCRIPTION: SEQ ID NO: 5477:
US-09-107-532A-5477

Query Match 1.0%; Score 8; DB 4; Length 332;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 130 EFINRQK 137
|||||
Db 137 EFINRQK 144

RESULT 16
US-09-134-001C-5116
Sequence 5116, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 5116
LENGTH: 490
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5116

Query Match 1.0%; Score 8; DB 4; Length 490;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 150 NDGAVALA 157
|||||
Db 72 NDGAVALA 79

RESULT 17
US-08-999-774A-10
Sequence 10, Application US/08999774A
Patent No. 6274312
GENERAL INFORMATION:
APPLICANT: Gish, Kurt C.
APPLICANT: Seghezzi, Wolfgang
APPLICANT: Shanahan, Frances
APPLICANT: Lees, Emma M.
APPLICANT: McClanahan, Terrill K.
TITLE OF INVENTION: Intracellular Regulatory Molecules;
TITLE OF INVENTION: Related Reagents
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: DNAX Research Institute
STREET: 901 California Avenue
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/999,774A
FILING DATE: 10-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/032,818
FILING DATE: 11-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ching, Edwin P.
REGISTRATION NUMBER: 34,090
REFERENCE/DOCKET NUMBER: DX0646
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650)852-9196
TELEFAX: (650)496-1200
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 503 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-999-774A-10

Query Match 1.0%; Score 8; DB 3; Length 503;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 357 PSPGPQA 364
|||||
DB 244 PSPGPQA 251

RESULT 18
US-08-221-583-57
Sequence 57, Application US/08221583
Patent No. 5486595
GENERAL INFORMATION:
APPLICANT: Heavner, George A.
TITLE OF INVENTION: Tumor Necrosis Factor Inhibitors
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5486595ris
STREET: One Liberty Place 46th Floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19403
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25:mdcmod.
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/221,583
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Deluca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: CCOR-0185
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-221-583-57

Query Match 0.9%; Score 7; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 QTPPEPS 358
|||||
DB 8 QTPPEPS 14

RESULT 19
US-08-221-583-58
Sequence 58, Application US/08221583
Patent No. 5486595
GENERAL INFORMATION:
APPLICANT: Heavner, George A.
TITLE OF INVENTION: Tumor Necrosis Factor Inhibitors
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5486595ris
STREET: One Liberty Place 46th Floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19403
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25:mdcmod.
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/221,583
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Deluca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: CCOR-0185
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-221-583-58

Query Match 0.9%; Score 7; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 QTPPEPS 358
|||||
DB 5 QTPPEPS 11

RESULT 20
US-08-221-583-59
Sequence 59, Application US/08221583
Patent No. 5486595
GENERAL INFORMATION:
APPLICANT: Heavner, George A.
TITLE OF INVENTION: Tumor Necrosis Factor Inhibitors
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5486595ris
STREET: One Liberty Place 46th Floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19403
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25:mdctcd.
CURRENT APPLICATION DATA: US/08/221,583
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Deluca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: CCR-0185
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-06-221-583-59

Query Match 0.9%; Score 7; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 QTRPEPS 358
|||||
DB 2 QTRPEPS 8

RESULT 21
PCT-US95-04018-57
Sequence 57, Application PC/TUS9504018
GENERAL INFORMATION:
APPLICANT: Heavner, George A.
APPLICANT: Kruszynski, Marian
APPLICANT: Mervic, Miljenko
APPLICANT: Weber, Robert W.
TITLE OF INVENTION: Tumor Necrosis Factor Inhibitors
NUMBER OF SEQUENCES: 76
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
ADDRESSEE: Norris
STREET: One Liberty Place 46th Floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19403
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04018
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/221,580
FILING DATE: 01-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/221,583
FILING DATE: 01-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/221,581
FILING DATE: 01-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Deluca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: CCR-0232
TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US95-04018-57

Query Match 0.9%; Score 7; DB 5; Length 15;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 QTRPEPS 358
|||||
DB 8 QTRPEPS 14

RESULT 22
PCT-US95-04018-58
Sequence 58, Application PC/TUS9504018
GENERAL INFORMATION:
APPLICANT: Heavner, George A.
APPLICANT: Kruszynski, Marian
APPLICANT: Mervic, Miljenko
APPLICANT: Weber, Robert W.
TITLE OF INVENTION: Tumor Necrosis Factor Inhibitors
NUMBER OF SEQUENCES: 76
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
ADDRESSEE: Norris
STREET: One Liberty Place 46th Floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19403
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04018
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/221,580
FILING DATE: 01-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/221,583
FILING DATE: 01-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/221,581
FILING DATE: 01-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Deluca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: CCR-0232
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US95-04018-58

Query Match 0.9%; Score 7; DB 5; Length 15;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 OPTPEPS 358
Db 5 OPTPEPS 11

RESULT 23
PCT-US95-04018-59
; Sequence 59, Application PC/TUS9504018
; GENERAL INFORMATION:
; APPLICANT: Heavner, George A.
; APPLICANT: Kruszynski, Marian
; APPLICANT: Mervic, Miljenko
; APPLICANT: Weber, Robert W.
; TITLE OF INVENTION: Tumor Necrosis Factor Inhibitors
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; ADDRESSEE: Norris
; STREET: One Liberty Place 46th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19403
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04018
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/221,580
; FILING DATE: 01-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/221,583
; FILING DATE: 01-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/221,581
; FILING DATE: 01-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: CCCR-0232
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; PCT-US95-04018-59

Query Match 0.9%; Score 7; DB 5; Length 15;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 OPTPEPS 358
Db 2 OPTPEPS 8

RESULT 24
US-09-020-116-4
; Sequence 4, Application US/09020116
; Patent No. 6084063
; GENERAL INFORMATION:
; APPLICANT: Vonakis, Becky
; APPLICANT: Metzger, Henry

; APPLICANT: Chen, Huaxian
; TITLE OF INVENTION: SIGNAL TRANSDUCTION INHIBITORS
; TITLE OF INVENTION: OF ALLERGIC REACTIONS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: SUITE 1200, 127 PEACHTREE STREET
; CITY: ATLANTA
; STATE: GA
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/020,116
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: MILLER, MARY L
; REGISTRATION NUMBER: 39,303
; REFERENCE/DOCKET NUMBER: 14014.0285
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 27 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-020-116-4

Query Match 0.9%; Score 7; DB 3; Length 27;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 557 GKDSLSLSD 563
Db 8 GKDSLSLSD 14

RESULT 25
US-09-608-902-4
; Sequence 4, Application US/09608902
; Patent No. 6353097
; GENERAL INFORMATION:
; APPLICANT: Vonakis, Becky
; APPLICANT: Metzger, Henry
; APPLICANT: Chen, Huaxian
; TITLE OF INVENTION: SIGNAL TRANSDUCTION INHIBITORS OF
; TITLE OF INVENTION: ALLERGIC REACTIONS
; FILE REFERENCE: 14014.0285U2
; CURRENT APPLICATION NUMBER: US/09/608,902
; CURRENT FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 09/020,116
; PRIOR FILING DATE: 1998-02-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 27
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: /No. 6353097e = synthetic con.

Query Match 0.9%; Score 7; DB 4; Length 27;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 557 GKDSLSD 563
Db 8 GKDSLSD 14

RESULT 26
US-09-020-116-2

; Sequence 2, Application US/09020116
; Patent No. 6084063
; GENERAL INFORMATION:
; APPLICANT: Vonakis, Becky
; APPLICANT: Metzger, Henry
; APPLICANT: Chen, Huaxian
; TITLE OF INVENTION: SIGNAL TRANSDUCTION INHIBITORS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: SUITE 1200, 127 PEACHTREE STREET
; CITY: ATLANTA
; STATE: GA
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/020,116
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: MILLER, MARY L
; REGISTRATION NUMBER: 39,303
; REFERENCE/DOCKET NUMBER: 14014.0285
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 45 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-020-116-2

Query Match 0.9%; Score 7; DB 3; Length 45;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 557 GKDSLSD 563
Db 8 GKDSLSD 14

RESULT 27
US-09-608-902-2

; Sequence 2, Application US/09608902
; Patent No. 6353097
; GENERAL INFORMATION:
; APPLICANT: Vonakis, Becky
; APPLICANT: Metzger, Henry
; APPLICANT: Chen, Huaxian
; TITLE OF INVENTION: SIGNAL TRANSDUCTION INHIBITORS OF

; TITLE OF INVENTION: ALLERGIC REACTIONS
; FILE REFERENCE: 14014.0285U2
; CURRENT APPLICATION NUMBER: US/09/608,902
; CURRENT FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 09/020,116
; PRIOR FILING DATE: 1998-02-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: /No. 6353097e = synthetic cons
; US-09-608-902-2

Query Match 0.9%; Score 7; DB 4; Length 45;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 557 GKDSLSD 563
Db 8 GKDSLSD 14

RESULT 28

US-09-171-646-1
; Sequence 1, Application US/09171646A
; Patent No. 6235516
; GENERAL INFORMATION:
; APPLICANT: Ghisalba, Oreste
; APPLICANT: Kittelmann, Mathias
; APPLICANT: Laumen, Kurt
; APPLICANT: Walser-Volken, Paula
; TITLE OF INVENTION: Biocatalysts with Amine Acylase Activity
; FILE REFERENCE: 4-20825/A/PCT
; CURRENT APPLICATION NUMBER: US/09/171,646A
; CURRENT FILING DATE: 1998-11-12
; EARLIER APPLICATION NUMBER: 96B10266.5
; EARLIER FILING DATE: 1996-04-25
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 50
; TYPE: PRT
; ORGANISM: Rhodococcus globerulus
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (41)..)
; OTHER INFORMATION: amino acid is uncertain
; US-09-171-646-1

Query Match 0.9%; Score 7; DB 3; Length 50;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 201 SASLAA 207
Db 9 SASLAA 15

RESULT 29

US-09-020-116-1
; Sequence 1, Application US/09020116
; Patent No. 6084063
; GENERAL INFORMATION:
; APPLICANT: Vonakis, Becky
; APPLICANT: Metzger, Henry
; APPLICANT: Chen, Huaxian
; TITLE OF INVENTION: SIGNAL TRANSDUCTION INHIBITORS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:

ADDRESSEE: NEEDLE & ROSENBERG, P.C.
STREET: SUITE 1200, 127 PEACHTREE STREET
CITY: ATLANTA
STATE: GA
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/020,116
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: MILLER, MARY L
REGISTRATION NUMBER: 39,303
REFERENCE/DOCKET NUMBER: 14014.0285
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 66 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-020-116-1

Query Match 0.9%; Score 7; DB 3; Length 66;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 557 GDSLSLSD 563
Db 8 GDSLSLSD 14

RESULT 30
US-09-608-902-1
Sequence 1, Application US/09608902
Patent No. 6353097
GENERAL INFORMATION:
APPLICANT: Vonakis, Becky
APPLICANT: Metzger, Henry
APPLICANT: Chen, Huaxian
TITLE OF INVENTION: SIGNAL TRANSDUCTION INHIBITORS OF
FILE REFERENCE: 14014.0285U2
CURRENT APPLICATION NUMBER: US/09/608,902
CURRENT FILING DATE: 2000-07-03
PRIOR APPLICATION NUMBER: 09/020,116
PRIOR FILING DATE: 1998-02-06
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 66
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: /No. 6353097e = synthetic cons
US-09-608-902-1

Query Match 0.9%; Score 7; DB 4; Length 66;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 557 GDSLSLSD 563

Db 8 GDSLSLSD 14

RESULT 31
US-09-328-352-6019
Sequence 6019, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 6019
LENGTH: 76
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-6019

Query Match 0.9%; Score 7; DB 4; Length 76;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 77 AIISEEL 83
Db 59 AIISEEL 65

RESULT 32
US-09-134-001C-4585
Sequence 4585, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4585
LENGTH: 117
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4585

Query Match 0.9%; Score 7; DB 4; Length 117;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 460 RNSDFQA 466
Db 11 RNSDFQA 17

RESULT 33
US-09-252-991A-28698
Sequence 28698, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 28698
LENGTH: 149
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28698

Query Match 0.9%; Score 7; DB 4; Length 149;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 203 SELAAE 209
Db 56 SELAAE 62

RESULT 34

US-08-917-456-2
Sequence 2, Application US/08917456
Patent No. 586368
GENERAL INFORMATION:
APPLICANT: Greenwood, Rebecca C.
APPLICANT: Gentry, Daniel R.
TITLE OF INVENTION: NOVEL GRE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dechert Price & Rhoads
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/917,456
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Dickinson, Todd Q
REGISTRATION NUMBER: 28,354
REFERENCE/DOCKET NUMBER: GM10046
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-994-2252
TELEFAX: 215-994-2222
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 158 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-917-456-2

Query Match 0.9%; Score 7; DB 2; Length 158;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 176 IIEDTGD 182
Db 77 IIEDTGD 83

RESULT 35

US-09-229-804-2
Sequence 2, Application US/09229804
Patent No. 6210674
GENERAL INFORMATION:
APPLICANT: Greenwood, Rebecca C.
APPLICANT: Gentry, Daniel R.
TITLE OF INVENTION: NOVEL GRE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dechert Price & Rhoads
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/229,804
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/917,456
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Dickinson, Todd Q
REGISTRATION NUMBER: 28,354
REFERENCE/DOCKET NUMBER: GM10046
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-994-2252
TELEFAX: 215-994-2222
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 158 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-229-804-2

Query Match 0.9%; Score 7; DB 3; Length 158;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 176 IIEDTGD 182
Db 77 IIEDTGD 83

RESULT 36

US-09-252-991A-20622
Sequence 20622, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 20622
LENGTH: 163
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-20622

Query Match 0.9%; Score 7; DB 4; Length 163;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 259 SQASQSN 265
DB 27 SQASQSN 33

RESULT 37

US-09-252-991A-30217
; Sequence 30217, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30217
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30217

Query Match 0.9%; Score 7; DB 4; Length 191;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 487 DLLAFLA 493
DB 174 DLLAFLA 180

RESULT 38

US-09-613-303-35
; Sequence 35, Application US/09613303
; Patent No. 6495347
; GENERAL INFORMATION:
; APPLICANT: Siegel, Marvin
; APPLICANT: Chu, N. Randall
; APPLICANT: Mizzen, Lee A.
; TITLE OF INVENTION: INDUCTION OF A TH1-LIKE RESPONSE IN VITRO
; FILE REFERENCE: 12071/002001
; CURRENT APPLICATION NUMBER: US/09/613,303
; PRIOR FILING DATE: 2000-07-10
; PRIOR APPLICATION NUMBER: US 60/143,757
; PRIOR FILING DATE: 1999-07-08
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 198
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fusion sequence
US-09-613-303-35

Query Match 0.9%; Score 7; DB 4; Length 198;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 GEXRIPL 611
DB 55 GEXRIPL 61

RESULT 39

US-08-916-901-3
; Sequence 3, Application US/08916901
; Patent No. 5892012
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: RAB PROTEINS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/916,901
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0367 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 201 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: LIVRIT04
; CLONE: 2514506
US-08-916-901-3

Query Match 0.9%; Score 7; DB 2; Length 201;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 365 PNLKIDS 371
DB 184 PNLKIDS 190

RESULT 40

US-08-916-901-8
; Sequence 8, Application US/08916901
; Patent No. 5892012
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: RAB PROTEINS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto

STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/916,901
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0367 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 201 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 57006
US-08-916-901-8

Query Match 0.9%; Score 7; DB 2; Length 201;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 365 PNLKIDS 371
184 PNLKIDS 190
Db

RESULT 41
US-09-154-602-3
Sequence 3, Application US/09154602
Patent No. 6300472
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Corley, Neil C.
APPLICANT: Shah, Purvi
TITLE OF INVENTION: RAB PROTEINS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/154,602
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/916,901
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0367 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 201 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: L1V7UT04
CLONE: 2514506
US-09-154-602-3

Query Match 0.9%; Score 7; DB 4; Length 201;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 365 PNLKIDS 371
184 PNLKIDS 190
Db

RESULT 42
US-09-154-602-8
Sequence 8, Application US/09154602
Patent No. 6300472
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Corley, Neil C.
APPLICANT: Shah, Purvi
TITLE OF INVENTION: RAB PROTEINS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/154,602
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/916,901
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0367 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 201 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 57006
US-09-154-602-8

Query Match 0.9%; Score 7; DB 4; Length 201;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;

	Matches	7; Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	365	PNLKIDS	371						
Db	184	PNLKIDS	190						

```

RESULT 43
US-08-908-332-5
; Sequence 5, Application US/08908332
; Patent No. 6013515
; GENERAL INFORMATION:
; APPLICANT: Xiao, Hua
; APPLICANT: Greenblatt, Jacka F.
; APPLICANT: Roeder, Robert G.
; TITLE OF INVENTION: COFACTORS FOR HIV-1 PROTEIN TAT AND
; TITLE OF INVENTION: METHODS OF USE THEREFOR
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; STREET: Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/908,332
; FILING DATE: 07-AUG-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-199
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-343-1684
; TELEFAX: 201-487-5800
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 226 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: E. coli
; US-08-908-332-5

Query Match 0.9%; Score 7; DB 3; Length 226;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0;

Qy 600 YNRVKG 606
|||
|||
Db 137 YNRVKG 143

RESULT 44
US-09-326-394-4
; Sequence 4, Application US/09326394
; Patent No. 6306820
; GENERAL INFORMATION:
; APPLICANT: Bendele, Alison M.
; APPLICANT: Sennello, Regina M.
; APPLICANT: Edwards, Carl K.
; TITLE OF INVENTION: COMBINATION THERAPY USING A TNF BINDING
; TITLE OF INVENTION: PROTEIN FOR TREATING TNF-MEDIATED DISEASES

```

```

NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen Inc.
STREET: 1840 DeHavilland Drive
CITY: Thousand Oaks
STATE: CA
COUNTRY: US
ZIP: 91320-1789

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/326,394
FILING DATE: 08-DEC-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/032,587
FILING DATE: 06-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/036,355
FILING DATE: 23-JAN-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/039,315
FILING DATE: 07-FEB-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/052,023
FILING DATE: 09-JUL-1997
ATTORNEY/AGENT INFORMATION:
NAME: Zindrick, Thomas K.
REGISTRATION NUMBER: 32,185
REFERENCE/DOCKET NUMBER: A-430D
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 235 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-326-394-4

Query Match 0.9%; Score 7; DB 4; Length 235;
Best Local Similarity 100.0%; Pctd. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 352 QPTPEPS 358
|||
Db 206 QPTPEPS 212

RESULT 45
US-09-580-235-2
Sequence 2, Application US/09580235
Patent No. 6433158
GENERAL INFORMATION:
APPLICANT: Petlic, Dean
TITLE OF INVENTION: Site Specific Protein Modification
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Janis C Henry
STREET: St University
CITY: Seattle
STATE: WA
COUNTRY: US
ZIP: 98101

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/580,235
FILING DATE:

```

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/102,530
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Henry, Janis C
REGISTRATION NUMBER: 34,347
REFERENCE/DOCKET NUMBER: 2637
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)470-4189
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 235 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-580-235-2

Query Match 0.9%; Score 7; DB 4; Length 235;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 352 QTPPEPS 358
|||||
Db 206 QTPPEPS 212

RESULT 46
US-09-580-235-4
Sequence 4, Application US/09580235
Patent No. 6433158
GENERAL INFORMATION:
APPLICANT: Pettit, Dean
TITLE OF INVENTION: Site Specific Protein Modification
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Janis C Henry
STREET: 51 University
CITY: Seattle
STATE: WA
COUNTRY: US
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/580,235
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/102,530
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Henry, Janis C
REGISTRATION NUMBER: 34,347
REFERENCE/DOCKET NUMBER: 2637
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)470-4189
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 235 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-580-235-4

Query Match 0.9%; Score 7; DB 4; Length 235;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 352 QTPPEPS 358

Db 206 QTPPEPS 212
|||||

RESULT 47
US-09-580-235-6
Sequence 6, Application US/09580235
Patent No. 6433158
GENERAL INFORMATION:
APPLICANT: Pettit, Dean
TITLE OF INVENTION: Site Specific Protein Modification
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Janis C Henry
STREET: 51 University
CITY: Seattle
STATE: WA
COUNTRY: US
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/580,235
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/102,530
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Henry, Janis C
REGISTRATION NUMBER: 34,347
REFERENCE/DOCKET NUMBER: 2637
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)470-4189
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 235 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-580-235-6

Query Match 0.9%; Score 7; DB 4; Length 235;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 352 QTPPEPS 358
|||||
Db 206 QTPPEPS 212

RESULT 48
US-09-580-235-8
Sequence 8, Application US/09580235
Patent No. 6433158
GENERAL INFORMATION:
APPLICANT: Pettit, Dean
TITLE OF INVENTION: Site Specific Protein Modification
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Janis C Henry
STREET: 51 University
CITY: Seattle
STATE: WA
COUNTRY: US
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/580,235
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/102,530
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Henry, Janis C
REGISTRATION NUMBER: 34,347
REFERENCE/DOCKET NUMBER: 2637
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)470-4189
INFORMATION FOR SEQ ID NO: 8
SEQUENCE CHARACTERISTICS:
LENGTH: 235 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-580-235-8

Query Match 0.9%; Score 7; DB 4; Length 235;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 OPTPEPS 358
Db 206 OPTPEPS 212

RESULT 49
US-09-580-181-2
Sequence 2, Application US/09580181
Patent No. 6441136
GENERAL INFORMATION:
APPLICANT: Pettit, Dean
TITLE OF INVENTION: Site Specific Protein Modification
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Janis C Henry
STREET: 51 University
CITY: Seattle
STATE: WA
COUNTRY: US
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/580,181
FILING DATE: 26-MAY-2000
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/102,530
FILING DATE: 22-JUN-1998
ATTORNEY/AGENT INFORMATION:
NAME: Henry, Janis C
REGISTRATION NUMBER: 34,347
REFERENCE/DOCKET NUMBER: 2637
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)470-4189
INFORMATION FOR SEQ ID NO: 2
SEQUENCE CHARACTERISTICS:
LENGTH: 235 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-580-181-2

Query Match 0.9%; Score 7; DB 4; Length 235;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 OPTPEPS 358
Db 206 OPTPEPS 212

RESULT 50
US-09-580-181-4
Sequence 4, Application US/09580181
Patent No. 6441136
GENERAL INFORMATION:
APPLICANT: Pettit, Dean
TITLE OF INVENTION: Site Specific Protein Modification
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Janis C Henry
STREET: 51 University
CITY: Seattle
STATE: WA
COUNTRY: US
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/580,181
FILING DATE: 26-MAY-2000
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/102,530
FILING DATE: 22-JUN-1998
ATTORNEY/AGENT INFORMATION:
NAME: Henry, Janis C
REGISTRATION NUMBER: 34,347
REFERENCE/DOCKET NUMBER: 2637
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)470-4189
INFORMATION FOR SEQ ID NO: 4
SEQUENCE CHARACTERISTICS:
LENGTH: 235 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-580-181-4

Query Match 0.9%; Score 7; DB 4; Length 235;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 OPTPEPS 358
Db 206 OPTPEPS 212

RESULT 51
US-09-580-181-6
Sequence 6, Application US/09580181
Patent No. 6441136
GENERAL INFORMATION:
APPLICANT: Pettit, Dean
TITLE OF INVENTION: Site Specific Protein Modification
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Janis C Henry
STREET: 51 University
CITY: Seattle
STATE: WA
COUNTRY: US
ZIP: 98101

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/580,181
FILING DATE: 26-MAY-2000
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/102,530
FILING DATE: 22-JUN-1998
ATTORNEY/AGENT INFORMATION:
NAME: Henry, Janis C
REGISTRATION NUMBER: 34,347
REFERENCE/DOCKET NUMBER: 2637
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)470-4189
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 235 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-580-181-6

Query Match 0.9%; Score 7; DB 4; Length 235;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 352 QPTPEPS 358
|||
Db 206 QPTPEPS 212

RESULT 52
US-09-580-181-8
Sequence 8, Application US/09580181
Patent No. 6441136
GENERAL INFORMATION:
APPLICANT: Pettit, Dean
TITLE OF INVENTION: Site Specific Protein Modification
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSER: Janis C Henry
STREET: 51 University
CITY: Seattle
STATE: WA
COUNTRY: US
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/580,181
FILING DATE: 26-MAY-2000
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/102,530
FILING DATE: 22-JUN-1998
ATTORNEY/AGENT INFORMATION:
NAME: Henry, Janis C
REGISTRATION NUMBER: 34,347
REFERENCE/DOCKET NUMBER: 2637
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)470-4189
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 235 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
US-09-580-181-8

Query Match 0.9%; Score 7; DB 4; Length 235;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 352 QPTPEPS 358
|||
Db 206 QPTPEPS 212

RESULT 53
US-09-102-530-2
Sequence 2, Application US/09102530
Patent No. 6451986
GENERAL INFORMATION:
APPLICANT: Pettit, Dean
TITLE OF INVENTION: Site Specific Protein Modification
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSER: Janis C Henry
STREET: 51 University
CITY: Seattle
STATE: WA
COUNTRY: US
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/102,530
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Henry, Janis C
REGISTRATION NUMBER: 34,347
REFERENCE/DOCKET NUMBER: 2637
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)470-4189
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 235 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-102-530-2

Query Match 0.9%; Score 7; DB 4; Length 235;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 352 QPTPEPS 358
|||
Db 206 QPTPEPS 212

RESULT 54
US-09-102-530-4
Sequence 4, Application US/09102530
Patent No. 6451986
GENERAL INFORMATION:
APPLICANT: Pettit, Dean
TITLE OF INVENTION: Site Specific Protein Modification
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSER: Janis C Henry
STREET: 51 University
CITY: Seattle
STATE: WA
COUNTRY: US

ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/102,530
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Henry, Janis C
REGISTRATION NUMBER: 34,347
REFERENCE/DOCKET NUMBER: 2637
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)470-4189
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 235 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-102-530-4

Query Match: 0.9%; Score 7; DB 4; Length 235;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 OPTPEPS 358
DB 206 OPTPEPS 212

RESULT 55
US-09-102-530-6
Sequence 6, Application US/09102530
Patent No. 6451986
GENERAL INFORMATION:
APPLICANT: Pettit, Dean
TITLE OF INVENTION: Site Specific Protein Modification
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Janis C Henry
STREET: 51 University
CITY: Seattle
STATE: WA
COUNTRY: US
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/102,530
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Henry, Janis C
REGISTRATION NUMBER: 34,347
REFERENCE/DOCKET NUMBER: 2637
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)470-4189
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 235 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-102-530-6

Query Match: 0.9%; Score 7; DB 4; Length 235;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 OPTPEPS 358
DB 206 OPTPEPS 212

RESULT 56
US-09-102-530-8
Sequence 8, Application US/09102530
Patent No. 6451986
GENERAL INFORMATION:
APPLICANT: Pettit, Dean
TITLE OF INVENTION: Site Specific Protein Modification
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Janis C Henry
STREET: 51 University
CITY: Seattle
STATE: WA
COUNTRY: US
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/102,530
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Henry, Janis C
REGISTRATION NUMBER: 34,347
REFERENCE/DOCKET NUMBER: 2637
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)470-4189
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 235 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-102-530-8

Query Match: 0.9%; Score 7; DB 4; Length 235;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 OPTPEPS 358
DB 206 OPTPEPS 212

RESULT 57
US-09-620-405B-487
Sequence 487, Application US/09620405B
Patent No. 6528054
GENERAL INFORMATION:
APPLICANT: Jiang, Yugu
APPLICANT: Dillon, David C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Xu, Jiaqun
APPLICANT: Harlocker, Susan L.
APPLICANT: Hepler, William T.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.470C8
CURRENT APPLICATION NUMBER: US/09/620,405B
CURRENT FILING DATE: 2000-07-20
NUMBER OF SEQ ID NOS: 495
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 487

LENGTH: 235
TYPE: PRT
ORGANISM: Homo sapiens
US-09-620-405B-487

Query Match 0.9%; Score 7; DB 4; Length 235;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 127 RTKEIN 133
Db 47 RTKEIN 53

RESULT 58
US-09-604-287A-487
Sequence 487, Application US/09604287A
Patent No. 6586572
GENERAL INFORMATION:

APPLICANT: Jiang, Yugu
APPLICANT: Dillon, Devin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Xu, Jianshun
APPLICANT: Harlocker, Susan L.
APPLICANT: Hepler, William T.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.470C7
CURRENT APPLICATION NUMBER: US/09/604,287A
CURRENT FILING DATE: 2000-06-22
NUMBER OF SEQ ID NOS: 489
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 487
LENGTH: 235
TYPE: PRT
ORGANISM: Homo sapiens
US-09-604-287A-487

Query Match 0.9%; Score 7; DB 4; Length 235;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 127 RTKEIN 133
Db 47 RTKEIN 53

RESULT 59
US-09-107-532A-6634
Sequence 6634, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 6634:
SEQUENCE CHARACTERISTICS:
LENGTH: 240 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1..240
SEQUENCE DESCRIPTION: SEQ ID NO: 6634:
US-09-107-532A-6634

Query Match 0.9%; Score 7; DB 4; Length 240;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 252 SNNSTN 258
Db 214 SNNSTN 220

RESULT 60
US-09-252-991A-29404
Sequence 29404, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 29404
LENGTH: 245
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29404

Query Match 0.9%; Score 7; DB 4; Length 245;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 150 NDGAVAL 156
Db 206 NDGAVAL 212

RESULT 61
US-09-579-845-10
Sequence 10, Application US/09579845
Patent No. 6537540
GENERAL INFORMATION:
APPLICANT: Burstein, Haim
APPLICANT: Stepan, Anthony M.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR LOWERING

;; TITLE OF INVENTION: THE LEVEL OF TUMOR NECROSIS FACTOR (TNF) IN TNF-ASSOCIATED
;; FILE REFERENCE: 226272004420
;; CURRENT APPLICATION NUMBER: US/09/579,845
;; CURRENT FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/150,688
;; NUMBER OF SEQ ID NOS: 20
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO: 10
;; LENGTH: 257
;; TYPE: PRT
;; ORGANISM: Homo sapien
US-09-579-845-10

Query Match 0.9%; Score 7; DB 4; Length 257;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 QPTPEPS 358
DB 228 QPTPEPS 234

RESULT 62
US-08-622-353-2
; Sequence 2, Application US/08622353
; Patent No. 5700925
; GENERAL INFORMATION:
; APPLICANT: Bishai, William R.
; APPLICANT: Young, Douglas B.
; APPLICANT: Zhang, Ying
; APPLICANT: Demaio, James
; TITLE OF INVENTION: A STATIONARY PHASE, STRESS RESPONSE
; TITLE OF INVENTION: SIGMA FACTOR FROM MYCOBACTERIUM TUBERCULOSIS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Banner & Allegretti, LTD
; STREET: 1001 G Street, eleventh floor
; CITY: NW
; STATE: Washington DC
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/622,353
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hoscheit, Dale
; REGISTRATION NUMBER: 19090
; REFERENCE/DOCKET NUMBER: 3181.51220
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 261 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Mycobacterium tuberculosis
US-08-622-353-2

Query Match 0.9%; Score 7; DB 1; Length 261;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 201 SASLELAA 207
DB 143 SASLELAA 149

RESULT 63
US-08-622-352A-2
; Sequence 2, Application US/08622352A
; Patent No. 5824546
; GENERAL INFORMATION:
; APPLICANT: Bishai, William R.
; APPLICANT: Demaio, James
; TITLE OF INVENTION: REGULATION OF A SIGMA FACTOR
; TITLE OF INVENTION: FROM MYCOBACTERIUM TUBERCULOSIS
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Pillsbury Madison & Sutro, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Word Perfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/622,352A
; FILING DATE: 27-MAR-1996
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 261 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Mycobacterium tuberculosis
US-08-622-352A-2

Query Match 0.9%; Score 7; DB 2; Length 261;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 201 SASLELAA 207
DB 143 SASLELAA 149

RESULT 64
US-08-826-390-2
; Sequence 2, Application US/08826390
; Patent No. 6004764
; GENERAL INFORMATION:
; APPLICANT: Bishai, William R.
; APPLICANT: Young, Douglas B.
; APPLICANT: Zhang, Ying
; APPLICANT: Demaio, James
; TITLE OF INVENTION: Stationary Phase, Stress Response
; TITLE OF INVENTION: Sigma Factor From Mycobacterium tuberculosis, and
; TITLE OF INVENTION: Regulation Thereof
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Pillsbury Madison & Sutro, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3918
; COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/826.390
; FILING DATE: 27-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/622,353
; FILING DATE: 27-MAR-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/622,352
; FILING DATE: 27-MAR-1996
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 261 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Mycobacterium tuberculosis
;
US-08-826-390-2

Query Match          0.9%; Score 7; DB 3; Length 261;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      201 SASIELAA 207
      |||||
Db      143 SASIELAA 149

RESULT 65
US-05-252-991A-21944
; Sequence 21944, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21944
; LENGTH: 298
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
;
US-09-252-991A-21944

Query Match          0.9%; Score 7; DB 4; Length 298;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      600 YNRVKGGE 606
      |||||
Db      210 YNRVKGGE 216

RESULT 66
US-09-252-991A-17702
; Sequence 17702, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17702
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
;
US-09-252-991A-17702

Query Match          0.9%; Score 7; DB 4; Length 305;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      152 GAVALAR 158
      |||||
Db      276 GAVALAR 282

RESULT 67
US-09-252-991A-31029
; Sequence 31029, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31029
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
;
US-09-252-991A-31029

Query Match          0.9%; Score 7; DB 4; Length 312;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      151 DGAVALA 157
      |||||
Db      283 DGAVALA 289

RESULT 68
US-09-252-991A-29807
; Sequence 29807, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29807
; LENGTH: 325
; TYPE: PRT
```

```

; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17702
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
;
US-09-252-991A-17702

Query Match          0.9%; Score 7; DB 4; Length 305;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      152 GAVALAR 158
      |||||
Db      276 GAVALAR 282

RESULT 67
US-09-252-991A-31029
; Sequence 31029, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31029
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
;
US-09-252-991A-31029

Query Match          0.9%; Score 7; DB 4; Length 312;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      151 DGAVALA 157
      |||||
Db      283 DGAVALA 289

RESULT 68
US-09-252-991A-29807
; Sequence 29807, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29807
; LENGTH: 325
; TYPE: PRT
```

ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29807

Query Match
Best Local Similarity 100.0%; Score 7; DB 4; Length 325;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 358 SPQPOPA 364
DB 179 SPQPOPA 185

RESULT 69
US-09-198-452A-745
Sequence 745, Application US/09198452A
Patent No. 6559294

GENERAL INFORMATION:
APPLICANT: Griffiths, R.
TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
TITLE OF INVENTION: and treatment of infection
FILE REFERENCE: 9710-003-999

CURRENT APPLICATION NUMBER: US/09/198,452A
CURRENT FILING DATE: 1998-11-24
NUMBER OF SEQ ID NOS: 6849

SEQ ID NO 745

LENGTH: 381

TYPE: PRT

ORGANISM: Chlamydia pneumoniae

FEATURE:
NAME/KEY: SITE

LOCATION: 1...381

OTHER INFORMATION: Xaa=unknown or other

QY 741 VTDSLSLK 747
DB 353 VTDSLSLK 359

Query Match
Best Local Similarity 100.0%; Score 7; DB 4; Length 381;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 70
US-09-459-749D-17

Sequence 17, Application US/09459749D
Patent No. 6464975

GENERAL INFORMATION:
APPLICANT: Millis, Albert J. T.

TITLE OF INVENTION: Compositions and Methods for Altering Cell Migration
FILE REFERENCE: 0794.016A

CURRENT APPLICATION NUMBER: US/09/459,749D
CURRENT FILING DATE: 1999-12-10

PRIOR APPLICATION NUMBER: 60/111,856
PRIOR FILING DATE: 1998-12-11

NUMBER OF SEQ ID NOS: 17

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 17

LENGTH: 383

TYPE: PRT

ORGANISM: Sus scrofa

US-09-459-749D-17

Query Match
Best Local Similarity 100.0%; Score 7; DB 4; Length 383;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 573 TREKGL 579
DB 288 TREKGL 294

RESULT 71
US-08-861-774E-82

Sequence 82, Application US/08861774E
Patent No. 6297007

GENERAL INFORMATION:
APPLICANT: Waters, Barbara

APPLICANT: Miao, Vivian

APPLICANT: Ho, Yap

APPLICANT: Tong, Seow

TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
BIOACTIVE MOLECULES

FILE REFERENCE: 9993-006
CURRENT APPLICATION NUMBER: US/08/861,774E

CURRENT FILING DATE: 1997-05-22
NUMBER OF SEQ ID NOS: 94

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 82

LENGTH: 390

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Clone ps10

QY 303 GVAVPHG 309
DB 11 GVAVPHG 17

Query Match
Best Local Similarity 100.0%; Score 7; DB 3; Length 390;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 72
US-08-311-731A-12

Sequence 12, Application US/08311731A
Patent No. 6583266

GENERAL INFORMATION:
APPLICANT: SMITH, DOUGLAS

APPLICANT: MAO, JEN-I

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR

TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 411

CORRESPONDENCE ADDRESS:
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.

STREET: 600 ATLANTIC AVENUE
CITY: BOSTON

STATE: MASSACHUSETTS
COUNTRY: USA

ZIP: 02210

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/311,731A

FILING DATE:
CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.

REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: C0044/7125

TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/720-3500

TELEFAX: 617/720-2441
INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:
LENGTH: 406 amino acids

TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: MYCOBACTERIUM TUBERCULOSIS
US-08-311-731A-12

Query Match 0.9%; Score 7; DB 4; Length 406;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 153 AVALARS 159
DB 84 AVALARS 90

RESULT 73
US-09-252-991A-22113
Sequence 22113, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 22113
LENGTH: 447
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22113

Query Match 0.9%; Score 7; DB 4; Length 447;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 359 PGPOPAP 365
DB 214 PGPOPAP 220

RESULT 74
US-08-385-229-2
Sequence 2, Application US/08385229
Patent No. 5605690
GENERAL INFORMATION:
APPLICANT: Jacobs, Cindy A.
TITLE OF INVENTION: Method of Treating TNF-Dependent
TITLE OF INVENTION: Inflammation Using Tumor Necrosis Factor Antagonists
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/385,229
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/946,236

FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Wight, Christopher L.
REGISTRATION NUMBER: 31,680
REFERENCE/DOCKET NUMBER: 2503
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 587-0606
TELEFAX: (206) 587-0430
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 461 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-385-229-2

Query Match 0.9%; Score 7; DB 1; Length 461;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 QTPPEPS 358
DB 228 QTPPEPS 234

RESULT 75
US-08-650-000-2
Sequence 2, Application US/08650000
Patent No. 5945397
GENERAL INFORMATION:
APPLICANT: Smith, Craig A.
APPLICANT: Goodwin, Raymond G.
APPLICANT: Beckmann, M. Patricia
TITLE OF INVENTION: Tumor Necrosis Factor Receptors
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/650,000
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/468,453
FILING DATE:
APPLICATION NUMBER: US/08/038,765
FILING DATE:
APPLICATION NUMBER: US 403,241
FILING DATE: 05-SEP-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 405,370
FILING DATE: 11-SEP-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 421,417
FILING DATE: 13-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 523,635
FILING DATE: 10-MAY-1990
ATTORNEY/AGENT INFORMATION:
NAME: Wight, Christopher L.
REGISTRATION NUMBER: 31,680
REFERENCE/DOCKET NUMBER: 2501-D
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 587-0430

TELEFAX: (206) 233-0644
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 461 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-650-000-2

Query Match 0.9%; Score 7; DB 2; Length 461;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 QTPERS 358
|||||
DB 228 QTPERS 234

RESULT 76
US-09-042-785A-7
; Sequence 7, Application US/09042785A
; Patent No. 6194151
; GENERAL INFORMATION:
; APPLICANT: Busfield, Samantha J
; TITLE OF INVENTION: NOVEL MOLECULES OF THE TNF RECEPTOR SUPERFAMILY
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/042,785A
; FILING DATE: 17-MAR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/938,896
; FILING DATE: 26-SEP-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragoras, Amy E
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: MEI-001CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 461 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: Internal
US-09-042-785A-7

Query Match 0.9%; Score 7; DB 3; Length 461;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 QTPERS 358
|||||
DB 228 QTPERS 234

RESULT 77
US-08-477-347-3
; Sequence 3, Application US/08477347

; Patent No. 6232446
; GENERAL INFORMATION:
; APPLICANT: WALLACH, David
; APPLICANT: BIGDA, Jacek
; APPLICANT: BELETSKY, Igor
; APPLICANT: METT, Igor
; TITLE OF INVENTION: TNF LIGANDS
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,347
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/115,685
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 106271
; FILING DATE: 08-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Townsend, G. Kevin
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: WALLACH=10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 461 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-477-347-3

Query Match 0.9%; Score 7; DB 3; Length 461;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 QTPERS 358
|||||
DB 228 QTPERS 234

RESULT 78
US-09-006-353A-4
; Sequence 4, Application US/09006353A
; Patent No. 6261801
; GENERAL INFORMATION:
; APPLICANT: MEI, YING-FEI
; APPLICANT: YU, GUO-LIANG
; APPLICANT: GENTZ, REINER
; APPLICANT: ROSEN, STEVEN
; TITLE OF INVENTION: TUMOR NECROSIS FACTOR RECEPTOR 5
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HUMAN GENOME SCIENCES, INC.
; STREET: 9410 KEY WEST AVENUE
; CITY: ROCKVILLE
; STATE: MD
; COUNTRY: US
; ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/006.353A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BROOKES, ANDERS A
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PF341
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 461 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-006-353A-4

Query Match 0.9%; Score 7; DB 3; Length 461;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 QTPPEPS 358
|||||
DB 228 QTPPEPS 234

RESULT 79
US-08-476-862-2
Sequence 2, Application US/08476862
Patent No. 6262239
GENERAL INFORMATION:
APPLICANT: WALLACH, David
APPLICANT: BIGDA, Jacek
APPLICANT: BELETSKY, Igor
APPLICANT: METT, Igor
APPLICANT: ENGELMANN, Hartmut
TITLE OF INVENTION: TNF INHIBITORS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/476.862
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 107267
FILING DATE: 12-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 94039
FILING DATE: 06-APR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 91229
FILING DATE: 06-AUG-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 90339

FILING DATE: 18-MAY-1989
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: WALLACH=12A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 461 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-476-862-2

Query Match 0.9%; Score 7; DB 3; Length 461;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 QTPPEPS 358
|||||
DB 228 QTPPEPS 234

RESULT 80
US-09-573-986-4
Sequence 4, Application US/09573986
Patent No. 6455040
GENERAL INFORMATION:
APPLICANT: Wei, Ying-Fei
APPLICANT: Ni, Jian
APPLICANT: Gentz, Reiner
APPLICANT: Ruben, Steven
TITLE OF INVENTION: Tumor Necrosis Factor Receptor 5
FILE REFERENCE: 1488.1280004
CURRENT APPLICATION NUMBER: US/09/573.986
CURRENT FILING DATE: 2000-05-18
NUMBER OF SEQ ID NOS: 27
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 4
LENGTH: 461
TYPE: PRT
ORGANISM: Homo sapiens
US-09-573-986-4

Query Match 0.9%; Score 7; DB 4; Length 461;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 QTPPEPS 358
|||||
DB 228 QTPPEPS 234

RESULT 81
US-08-406-824A-2
Sequence 2, Application US/08406824A
Patent No. 6541610
GENERAL INFORMATION:
APPLICANT: SMITH, Craig A.
TITLE OF INVENTION: TUMOR NECROSIS FACTOR-ALPHA AND BETA-RECEPTORS
FILE REFERENCE: A-71592
CURRENT APPLICATION NUMBER: US/08/406.824A
CURRENT FILING DATE: 1995-03-20
PRIOR APPLICATION NUMBER: US 08/255,849
PRIOR FILING DATE: 1994-06-08
PRIOR APPLICATION NUMBER: US 07/860,710
PRIOR FILING DATE: 1992-03-30
PRIOR APPLICATION NUMBER: US 07/523,635
PRIOR FILING DATE: 1990-05-10
PRIOR APPLICATION NUMBER: US 07/421,417
PRIOR FILING DATE: 1989-10-13

PRIOR APPLICATION NUMBER: US 07/405,370
 PRIOR FILING DATE: 1989-09-11
 PRIOR APPLICATION NUMBER: US 07/403,241
 PRIOR FILING DATE: 1989-09-05
 NUMBER OF SEQ ID NOS: 29
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 2
 LENGTH: 461
 TYPE: prt
 ORGANISM: Homo sapiens
 US-08-406-824A-2

Query Match 0.9%; Score 7; DB 4; Length 461;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 OPTPEPS 358
 DB 228 OPTPEPS 234

RESULT 82
 US-09-800-909-2
 Sequence 2, Application US/09800909
 Patent No. 655511
 GENERAL INFORMATION:
 APPLICANT: WALLACH, David
 APPLICANT: BIGDA, Jacek
 APPLICANT: BELETSKY, Igor
 APPLICANT: MEYER, Igor
 APPLICANT: ENGELMANN, Hartmut
 TITLE OF INVENTION: TNF INHIBITORS
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: BROWDY AND NETMARK
 STREET: 419 Seventh Street, N.W.
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20004
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/800,909
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/476,862
 FILING DATE:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: IL 94039
 FILING DATE: 06-APR-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: IL 91229
 FILING DATE: 06-AUG-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: IL 90339
 FILING DATE: 18-MAY-1989
 ATTORNEY/AGENT INFORMATION:
 NAME: BROWDY, Roger L.
 REGISTRATION NUMBER: 25,618
 REFERENCE/DOCKET NUMBER: WALLACH=12A
 TELEPHONE: 202-628-5197
 TELEFAX: 202-737-3528
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 461 amino acids
 TYPE: amino acid
 TOPOLOGY: linear

MOLECULE TYPE: protein
 US-09-800-909-2

Query Match 0.9%; Score 7; DB 4; Length 461;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 OPTPEPS 358
 DB 228 OPTPEPS 234

RESULT 83
 US-09-758-124-2
 Sequence 2, Application US/09758124
 Patent No. 6572852
 GENERAL INFORMATION:
 APPLICANT: SMITH, Craig A.
 APPLICANT: GOODWIN, Raymond G.
 APPLICANT: BECKMANN, M. Patricia
 TITLE OF INVENTION: TUMOR NECROSIS FACTOR-ALPHA AND -BETA RECEPTORS
 FILE REFERENCE: A7895
 CURRENT APPLICATION NUMBER: US/09/758,124
 PRIOR FILING DATE: 2001-01-12
 PRIOR APPLICATION NUMBER: 08/953,268
 PRIOR FILING DATE: 1997-10-17
 PRIOR APPLICATION NUMBER: 08/555,629
 PRIOR FILING DATE: 1995-11-09
 PRIOR APPLICATION NUMBER: 08/468,453
 PRIOR FILING DATE: 1995-06-06
 PRIOR APPLICATION NUMBER: 08/038,765
 PRIOR FILING DATE: 1993-03-13
 PRIOR APPLICATION NUMBER: 07/523,635
 PRIOR FILING DATE: 1990-05-10
 PRIOR APPLICATION NUMBER: 07/421,417
 PRIOR FILING DATE: 1989-10-13
 PRIOR APPLICATION NUMBER: 07/405,370
 PRIOR FILING DATE: 1989-09-11
 PRIOR APPLICATION NUMBER: 07/403,241
 PRIOR FILING DATE: 1989-09-05
 NUMBER OF SEQ ID NOS: 4
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 2
 LENGTH: 461
 TYPE: prt
 ORGANISM: Homo sapiens
 US-09-758-124-2

Query Match 0.9%; Score 7; DB 4; Length 461;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 OPTPEPS 358
 DB 228 OPTPEPS 234

RESULT 84
 5395760-2
 Patent No. 5395760
 APPLICANT: SMITH, CRAIG A., GOODWIN, RAYMOND G., BECKMANN, M. PATRICIA
 TITLE OF INVENTION: DNA ENCODING TUMOR NECROSIS FACTOR-a AND -B RECEPTORS
 NUMBER OF SEQUENCES: 17
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/523,635
 FILING DATE: 10-MAY-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 421,417
 FILING DATE: 13-OCT-1989
 APPLICATION NUMBER: 405,370
 FILING DATE: 11-SEP-1989

APPLICATION NUMBER: 403,241
FILING DATE: 05-SEP-1989
SEQ ID NO. 2:
LENGTH: 461
5395760-2

Query Match 0.9%; Score 7; DB 6; Length 461;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 352 QPTPEPS 358
|||||
Db 228 QPTPEPS 234

RESULT 65
US-09-328-352-4742
Sequence 4742, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 4742
LENGTH: 462
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-4742

Query Match 0.9%; Score 7; DB 4; Length 462;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 415 LSKOESV 421
|||||
Db 42 LSKOESV 48

RESULT 86
US-09-252-991A-17797
Sequence 17797, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 17797
LENGTH: 485
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17797

Query Match 0.9%; Score 7; DB 4; Length 485;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 357 PSPGPOP 363
|||||
Db 2 PSPGPOP 8

RESULT 87
US-08-243-010-1
Sequence 1, Application US/08243010
Patent No. 5639597

GENERAL INFORMATION:
APPLICANT: Lauffer, Leander
APPLICANT: Zettlmeissel, Gerd
APPLICANT: Oquendo, Patricia
TITLE OF INVENTION: Cell-free Receptor Binding Assays, The
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/243,010
FILING DATE: 13-MAY-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/798,564
FILING DATE: 26-NOV-1991
APPLICATION NUMBER: DE P 40 37 837.3
FILING DATE: 28-NOV-1990

ATTORNEY/AGENT INFORMATION:
NAME: Einaudi, Carol P.
REGISTRATION NUMBER: 32,220
REFERENCE/DOCKET NUMBER: 02481-1132-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-243-010-1

Query Match 0.9%; Score 7; DB 1; Length 486;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 352 QPTPEPS 358
|||||
Db 228 QPTPEPS 234

RESULT 88
US-09-252-991A-27183
Sequence 27183, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 27183
 LENGTH: 498
 TYPE: PRT
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-27183

Query Match
 Best Local Similarity 100.0%; Score 7; DB 4; Length 498;
 Pred. No. 2.5e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 DEDIVNE 100
 DB 406 DEDIVNE 412

RESULT 89
 US-09-620-405B-478
 Sequence 478, Application US/09620405B
 Patent No. 6528054
 GENERAL INFORMATION:
 APPLICANT: Jiang, Yugu
 APPLICANT: Dillon, Davin C.
 APPLICANT: Mitcham, Jennifer L.
 APPLICANT: Xu, Jiaqichun
 APPLICANT: Harlocker, Susan L.
 APPLICANT: Hepler, William T.
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER
 FILE REFERENCE: 210121.470C8
 CURRENT APPLICATION NUMBER: US/09/620,405B
 CURRENT FILING DATE: 2000-07-20
 NUMBER OF SEQ ID NOS: 495
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 478
 LENGTH: 505
 TYPE: PRT
 ORGANISM: Homo sapien
 US-09-620-405B-478

Query Match
 Best Local Similarity 100.0%; Score 7; DB 4; Length 505;
 Pred. No. 2.5e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 RTKEIN 133
 DB 317 RTKEIN 323

RESULT 90
 US-09-620-405B-485
 Sequence 485, Application US/09620405B
 Patent No. 6528054
 GENERAL INFORMATION:
 APPLICANT: Jiang, Yugu
 APPLICANT: Dillon, Davin C.
 APPLICANT: Mitcham, Jennifer L.
 APPLICANT: Xu, Jiaqichun
 APPLICANT: Harlocker, Susan L.
 APPLICANT: Hepler, William T.
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER
 FILE REFERENCE: 210121.470C8
 CURRENT APPLICATION NUMBER: US/09/620,405B
 CURRENT FILING DATE: 2000-07-20
 NUMBER OF SEQ ID NOS: 495
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 485
 LENGTH: 505
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-620-405B-485

Query Match 0.9%; Score 7; DB 4; Length 505;

Best Local Similarity 100.0%; Pred. No. 2.5e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 RTKEIN 133
 DB 317 RTKEIN 323

RESULT 91
 US-09-604-287A-478
 Sequence 478, Application US/09604287A
 Patent No. 6586572
 GENERAL INFORMATION:
 APPLICANT: Jiang, Yugu
 APPLICANT: Dillon, Davin C.
 APPLICANT: Mitcham, Jennifer L.
 APPLICANT: Xu, Jiaqichun
 APPLICANT: Harlocker, Susan L.
 APPLICANT: Hepler, William T.
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER
 FILE REFERENCE: 210121.470C7
 CURRENT APPLICATION NUMBER: US/09/604,287A
 CURRENT FILING DATE: 2000-06-22
 NUMBER OF SEQ ID NOS: 489
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 478
 LENGTH: 505
 TYPE: PRT
 ORGANISM: Homo sapien
 US-09-604-287A-478

Query Match
 Best Local Similarity 100.0%; Score 7; DB 4; Length 505;
 Pred. No. 2.5e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 RTKEIN 133
 DB 317 RTKEIN 323

RESULT 92
 US-09-604-287A-485
 Sequence 485, Application US/09604287A
 Patent No. 6586572
 GENERAL INFORMATION:
 APPLICANT: Jiang, Yugu
 APPLICANT: Dillon, Davin C.
 APPLICANT: Mitcham, Jennifer L.
 APPLICANT: Xu, Jiaqichun
 APPLICANT: Harlocker, Susan L.
 APPLICANT: Hepler, William T.
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER
 FILE REFERENCE: 210121.470C7
 CURRENT APPLICATION NUMBER: US/09/604,287A
 CURRENT FILING DATE: 2000-06-22
 NUMBER OF SEQ ID NOS: 489
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 485
 LENGTH: 505
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-604-287A-485

Query Match
 Best Local Similarity 100.0%; Score 7; DB 4; Length 505;
 Pred. No. 2.5e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 RTKEIN 133
 DB 317 RTKEIN 323

```
RESULT 93
US-09-252-991A-19121
; Sequence 19121, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19121
; LENGTH: 506
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19121

Query Match      0.9%; Score 7; DB 4; Length 506;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      205 LAAAEAF 211
Db      104 LAAAEAF 110

RESULT 94
US-09-991-677-4
; Sequence 4, Application US/08991677A
; Patent No. 6252135
; GENERAL INFORMATION:
; APPLICANT: Chiang, Vincent L.
; APPLICANT: Carraway, Daniel T.
; APPLICANT: Smeltzer, Richard H.
; TITLE OF INVENTION: Production of Syringyl Lignin in Gymnosperms
; FILE REFERENCE: 50617
; CURRENT APPLICATION NUMBER: US/08/991,677A
; CURRENT FILING DATE: 1997-12-16
; EARLIER APPLICATION NUMBER: US 60/033,381
; EARLIER FILING DATE: 1996-12-16
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Liquidambar styraciflua
US-08-991-677-4

Query Match      0.9%; Score 7; DB 3; Length 511;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      485 VDDILAF 491
Db      270 VDDILAF 276

RESULT 95
US-08-426-509A-16
; Sequence 16, Application US/08426509A
; Patent No. 6326469
; GENERAL INFORMATION:
; APPLICANT: Ulirich, Axel
; APPLICANT: Gishizeky, Mikhail
; APPLICANT: Sures, Irman G.
; TITLE OF INVENTION: NOVEL MEGAKARYOCYTIC PROTEIN
; TITLE OF INVENTION: TYROSINE KINASES
```

```
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESS: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York,
STATE: NY
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/426,509A
FILING DATE: 21-APR-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/232,545
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7683-0074-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 512 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: No. 6326469e
US-08-426-509A-16

Query Match      0.9%; Score 7; DB 4; Length 512;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      557 GKDSLSD 563
Db      8 GKDSLSD 14

RESULT 96
US-08-232-545-16
; Sequence 16, Application US/08232545
; Patent No. 6506578
; GENERAL INFORMATION:
; APPLICANT: Ulirich, Axel
; APPLICANT: Gishizeky, Mikhail
; APPLICANT: Sures, Irman G.
; TITLE OF INVENTION: No. 6506578e1 Megakaryocytic Protein Tyrosine
; TITLE OF INVENTION: Kinases
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESS: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,545
; FILING DATE: 22-APR-1994
; CLASSIFICATION: 435
```

ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7683-050
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)790-9090
TELEFAX: (212)869-9741
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 512 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-232-545-16

Query Match
Best Local Similarity 100.0%; Score 7; DB 4; Length 512;
Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 557 GKDSLSD 563
|||||
DB 8 GKDSLSD 14

RESULT 97
PCT-US95-05008-16
Sequence 16, Application PC/TUS9505008
GENERAL INFORMATION:
APPLICANT: Sugen, Inc.
APPLICANT: 515 Galveston Drive
APPLICANT: Redwood City, California 94063-4720
APPLICANT: United States of America
APPLICANT: Wissenschaften E.V.
APPLICANT: Hofgarten Str. 2
APPLICANT: Munchen 80539
APPLICANT: Germany
TITLE OF INVENTION: Novel Megakaryocytic Protein Tyrosine
TITLE OF INVENTION: Kinases
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/05008
FILING DATE: 24-APR-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/232,545
FILING DATE: 22-APR-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7683-074
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)790-9090
TELEFAX: (212)869-9741
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 512 amino acids
TYPE: amino acid

STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
PCT-US95-05008-16

Query Match
Best Local Similarity 100.0%; Score 7; DB 5; Length 512;
Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 557 GKDSLSD 563
|||||
DB 8 GKDSLSD 14

RESULT 98
US-09-564-808-6
Sequence 6, Application US/09564808
Patent No. 6501004
GENERAL INFORMATION:
APPLICANT: Selvaraj, Gopalan
APPLICANT: Nair, Ramesh B
APPLICANT: Joy, IV, Richard W
APPLICANT: Keller, Wilfred A
APPLICANT: Datta, Raju S
TITLE OF INVENTION: Transgenic Reduction of Sinapine in Crucifera
FILE REFERENCE: 44618 Sequence Listing
Patent No. 6501004
CURRENT APPLICATION NUMBER: US/09/564,808
CURRENT FILING DATE: 2000-05-05
PRIOR APPLICATION NUMBER: 60/132,800
PRIOR FILING DATE: 1999-05-06
PRIOR APPLICATION NUMBER: 2,270,417
PRIOR FILING DATE: 1999-05-06
NUMBER OF SEQ ID NOS: 22
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 6
LENGTH: 513
TYPE: PRT
ORGANISM: Brassica napus
US-09-564-808-6

Query Match
Best Local Similarity 100.0%; Score 7; DB 4; Length 513;
Pred. No. 2.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 485 VDDLIAF 491
|||||
DB 271 VDDLIAF 277

RESULT 99
US-08-385-229-4
Sequence 4, Application US/08385229
Patent No. 5605690
GENERAL INFORMATION:
APPLICANT: Jacobs, Cindy A.
APPLICANT: Smith, Craig A.
TITLE OF INVENTION: Method of Treating TNF-Dependent
TITLE OF INVENTION: Inflammation Using Tumor Necrosis Factor Antagonists
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/385,229
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/07/946,236
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Wight, Christopher L.
 REGISTRATION NUMBER: 31,680
 REFERENCE/DOCKET NUMBER: 2503
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 587-0606
 TELEFAX: (206) 587-0430
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 518 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-385-229-4

Query Match 0.9%; Score 7; DB 1; Length 518;
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 352 OPTPEPS 358
 |||||
 Db 257 OPTPEPS 263

RESULT 100
 US-09-579-845-1
 ; Sequence 1, Application US/09579845
 ; Patent No. 6537540
 ; GENERAL INFORMATION:
 ; APPLICANT: Birstein, Haim
 ; APPLICANT: Stepan, Anthony M.
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR LOWERING
 ; TITLE OF INVENTION: THE LEVEL OF TUMOR NECROSIS FACTOR (TNF) IN TNF-ASSOCIATED
 ; TITLE OF INVENTION: DISORDERS
 ; FILE REFERENCE: 226272004420
 ; CURRENT APPLICATION NUMBER: US/09/579,845
 ; CURRENT FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/150,688
 ; PRIOR FILING DATE: 1999-05-28
 ; NUMBER OF SEQ ID NOS: 20
 ; SOFTWARE: fastseq for windows Version 4.0
 ; SEQ ID NO: 1
 ; LENGTH: 518
 ; TYPE: PRT
 ; ORGANISM: Homo sapien
 US-09-579-845-1

Query Match 0.9%; Score 7; DB 4; Length 518;
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 352 OPTPEPS 358
 |||||
 Db 257 OPTPEPS 263

Search completed: November 14, 2003, 11:04:05
 Job time : 122 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 11:02:11 ; Search time 69 Seconds
(without alignments)
2106.049 Million cell updates/sec

Title: US-09-765-271-56

Perfect score: 796

Sequence: 1 SYELGLYQARTVKNRVS.....KLIALIKSNPSVSKKIN 796

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 666188 seqs, 182559486 residues

Word size : 0

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 200 summaries

Database : Published Applications AA.*

1: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubppa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubppa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubppa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubppa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubppa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubppa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubppa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubppa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubppa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubppa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubppa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubppa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubppa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubppa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	795	99.9	796	9	US-09-765-272-56
2	80	10.1	840	11	US-09-884-465A-7
3	60	7.5	826	11	US-09-769-787-184
4	60	7.5	838	11	US-09-884-465A-8
5	51	6.4	753	9	US-09-765-272-66
6	32	4.0	42	11	US-09-884-465A-12
7	32	4.0	447	9	US-09-765-272-182
8	32	4.0	484	11	US-09-769-787-38
9	32	4.0	484	12	US-09-744A-24
10	32	4.0	840	11	US-09-884-465A-10
11	32	4.0	1039	11	US-09-884-465A-6
12	27	3.4	1139	11	US-09-884-465A-380
13	27	3.4	1378	11	US-09-884-465A-378
14	24	3.0	999	11	US-09-884-465A-376
15	24	3.0	999	11	US-09-884-465A-377

16	24	3.0	1126	11	US-09-884-465A-383	Sequence 383, App
17	24	3.0	1238	11	US-09-884-465A-381	Sequence 381, App
18	24	3.0	1365	11	US-09-884-465A-382	Sequence 382, App
19	23	2.9	94	11	US-09-884-465A-11	Sequence 11, Appl
20	17	2.1	381	12	US-09-769-736-24	Sequence 24, Appl
21	17	2.1	793	11	US-09-252-088-15	Sequence 15, Appl
22	17	2.1	822	12	US-09-769-726-18	Sequence 18, Appl
23	12	1.5	715	11	US-09-252-088-16	Sequence 16, Appl
24	12	1.5	913	11	US-09-884-465A-384	Sequence 384, App
25	12	1.5	1152	11	US-09-884-465A-379	Sequence 379, App
26	11	1.4	485	12	US-09-769-736-12	Sequence 72, Appl
27	9	1.1	272	11	US-09-884-465A-300	Sequence 300, App
28	8	1.0	86	11	US-09-884-465A-22	Sequence 22, Appl
29	8	1.0	259	11	US-09-884-465A-374	Sequence 374, App
30	8	1.0	270	11	US-09-884-465A-302	Sequence 302, App
31	8	1.0	272	11	US-09-884-465A-258	Sequence 258, App
32	8	1.0	272	11	US-09-884-465A-293	Sequence 293, App
33	8	1.0	400	15	US-10-156-761-13480	Sequence 13480, A
34	8	1.0	482	9	US-09-815-242-12564	Sequence 5401, Ap
35	8	1.0	488	9	US-09-815-242-12564	Sequence 12564, A
36	8	1.0	894	11	US-09-884-465A-336	Sequence 336, App
37	8	1.0	895	11	US-09-884-465A-344	Sequence 344, App
38	8	1.0	900	11	US-09-884-465A-335	Sequence 335, App
39	8	1.0	901	11	US-09-884-465A-343	Sequence 343, App
40	8	1.0	906	11	US-09-884-465A-332	Sequence 332, App
41	8	1.0	906	11	US-09-884-465A-369	Sequence 369, App
42	8	1.0	906	11	US-09-884-465A-370	Sequence 370, App
43	8	1.0	906	11	US-09-884-465A-371	Sequence 371, App
44	8	1.0	906	11	US-09-884-465A-372	Sequence 372, App
45	8	1.0	906	11	US-09-884-465A-373	Sequence 373, App
46	8	1.0	1104	15	US-10-128-714-3262	Sequence 3262, Ap
47	8	1.0	1289	10	US-09-712-333-259	Sequence 259, App
48	8	1.0	1353	15	US-10-128-714-8262	Sequence 8262, Ap
49	7	0.9	41	11	US-09-884-465A-23	Sequence 23, Appl
50	7	0.9	49	9	US-09-864-761-38022	Sequence 38022, A
51	7	0.9	81	15	US-10-106-658-5438	Sequence 5438, Ap
52	7	0.9	91	11	US-09-764-891-3850	Sequence 3850, Ap
53	7	0.9	97	15	US-10-103-313-297	Sequence 297, App
54	7	0.9	97	15	US-10-103-313-467	Sequence 467, App
55	7	0.9	100	10	US-09-712-363-272	Sequence 272, App
56	7	0.9	115	12	US-10-252-945-57	Sequence 57, Appl
57	7	0.9	139	15	US-10-156-761-7659	Sequence 7659, Appl
58	7	0.9	143	11	US-09-820-843A-94	Sequence 94, Appl
59	7	0.9	158	9	US-09-817-414-2	Sequence 2, Appl
60	7	0.9	158	15	US-10-043-467-355	Sequence 355, App
61	7	0.9	166	10	US-09-738-626-6669	Sequence 6669, Ap
62	7	0.9	169	10	US-09-882-529-5	Sequence 5, Appl
63	7	0.9	170	10	US-09-882-529-6	Sequence 6, Appl
64	7	0.9	170	10	US-09-882-529-8	Sequence 8, Appl
65	7	0.9	178	15	US-10-156-761-9896	Sequence 9896, Ap
66	7	0.9	198	15	US-10-267-311-35	Sequence 35, Appl
67	7	0.9	201	10	US-09-967-736-3	Sequence 3, Appl
68	7	0.9	209	10	US-09-967-736-8	Sequence 8, Appl
69	7	0.9	217	9	US-10-156-761-11020	Sequence 11020, A
70	7	0.9	224	15	US-09-925-301-1048	Sequence 1048, Ap
71	7	0.9	235	9	US-10-102-806-466	Sequence 466, App
72	7	0.9	235	9	US-09-604-287A-487	Sequence 487, App
73	7	0.9	235	10	US-09-907-263-4	Sequence 4, Appl
74	7	0.9	235	11	US-09-882-735-16	Sequence 16, Appl
75	7	0.9	235	12	US-10-124-805-487	Sequence 487, App
76	7	0.9	235	12	US-10-436-826-75	Sequence 75, Appl
77	7	0.9	235	14	US-10-007-805-487	Sequence 487, App
78	7	0.9	235	15	US-10-076-652-487	Sequence 487, App
79	7	0.9	235	15	US-10-243-230-2	Sequence 2, Appl
80	7	0.9	235	15	US-10-243-230-4	Sequence 4, Appl
81	7	0.9	235	15	US-10-243-230-6	Sequence 6, Appl
82	7	0.9	235	15	US-10-243-230-8	Sequence 8, Appl
83	7	0.9	247	15	US-10-156-761-12784	Sequence 12784, A
84	7	0.9	248	12	US-10-032-585-7603	Sequence 7603, Ap
85	7	0.9	257	15	US-10-313-852-10	Sequence 10, Appl
86	7	0.9	257	15	US-10-314-033-10	Sequence 10, Appl
87	7	0.9	264	15	US-10-156-761-11174	Sequence 11174, A
88	7	0.9	266	11	US-09-884-465A-304	Sequence 304, App

89	7	0.9	268	11	US-09-884-465A-303	Sequence 303, App
90	7	0.9	268	11	US-09-884-465A-303	Sequence 375, App
91	7	0.9	272	11	US-09-884-465A-295	Sequence 294, App
92	7	0.9	272	11	US-09-884-465A-295	Sequence 295, App
93	7	0.9	272	11	US-09-884-465A-296	Sequence 296, App
94	7	0.9	272	11	US-09-884-465A-298	Sequence 298, App
95	7	0.9	272	11	US-09-884-465A-298	Sequence 299, App
96	7	0.9	272	11	US-09-884-465A-299	Sequence 301, App
97	7	0.9	285	12	US-09-884-465A-301	Sequence 310, App
98	7	0.9	285	12	US-09-884-465A-301	Sequence 311, App
99	7	0.9	293	12	US-10-032-585-7197	Sequence 7197, App
100	7	0.9	311	10	US-09-884-055-113	Sequence 113, App
101	7	0.9	311	11	US-09-884-291-113	Sequence 113, App
102	7	0.9	311	12	US-10-219-810-49	Sequence 49, App
103	7	0.9	311	12	US-10-044-643-16	Sequence 16, App
104	7	0.9	311	12	US-10-044-643-18	Sequence 18, App
105	7	0.9	311	12	US-10-044-643-20	Sequence 20, App
106	7	0.9	311	15	US-10-220-382-10	Sequence 10, App
107	7	0.9	336	10	US-09-738-626-4819	Sequence 4819, App
108	7	0.9	351	15	US-10-146-772-78	Sequence 78, App
109	7	0.9	356	15	US-10-220-382-9	Sequence 9, App
110	7	0.9	374	15	US-10-268-441-2	Sequence 2, App
111	7	0.9	378	15	US-10-103-313-334	Sequence 434, App
112	7	0.9	379	12	US-09-841-260-105	Sequence 105, App
113	7	0.9	379	14	US-10-007-693-105	Sequence 105, App
114	7	0.9	383	10	US-09-459-749D-17	Sequence 17, App
115	7	0.9	390	10	US-09-924-256A-82	Sequence 82, App
116	7	0.9	410	12	US-10-238-075-1228	Sequence 1228, App
117	7	0.9	424	15	US-10-207-655-192	Sequence 192, App
118	7	0.9	434	12	US-10-032-585-7690	Sequence 7690, App
119	7	0.9	436	9	US-09-925-301-928	Sequence 928, App
120	7	0.9	449	15	US-10-156-761-13001	Sequence 13001, A
121	7	0.9	450	10	US-09-768-779A-3	Sequence 3, App
122	7	0.9	450	10	US-10-291-480-3	Sequence 3, App
123	7	0.9	459	12	US-10-032-585-7777	Sequence 7777, App
124	7	0.9	459	15	US-10-102-806-720	Sequence 720, App
125	7	0.9	461	9	US-09-800-909-2	Sequence 2, App
126	7	0.9	461	9	US-09-826-212-4	Sequence 4, App
127	7	0.9	461	9	US-09-758-124-2	Sequence 2, App
128	7	0.9	461	9	US-09-896-096A-17	Sequence 17, App
129	7	0.9	461	9	US-09-894-924-17	Sequence 17, App
130	7	0.9	461	10	US-09-840-707A-17	Sequence 17, App
131	7	0.9	461	10	US-09-800-908-3	Sequence 3, App
132	7	0.9	461	10	US-09-935-727-6	Sequence 6, App
133	7	0.9	461	11	US-09-902-176A-50	Sequence 50, App
134	7	0.9	461	11	US-09-902-176A-52	Sequence 52, App
135	7	0.9	461	11	US-09-902-176A-54	Sequence 54, App
136	7	0.9	461	12	US-10-420-785-2	Sequence 2, App
137	7	0.9	461	14	US-10-164-592-3	Sequence 3, App
138	7	0.9	461	15	US-10-252-408-2	Sequence 2, App
139	7	0.9	461	15	US-10-046-433-6	Sequence 6, App
140	7	0.9	461	15	US-10-038-557A-17	Sequence 17, App
141	7	0.9	461	15	US-10-166-643-4	Sequence 4, App
142	7	0.9	490	12	US-10-363-427-4	Sequence 4, App
143	7	0.9	493	9	US-09-815-242-12052	Sequence 12052, A
144	7	0.9	496	15	US-10-011-585A-199	Sequence 199, App
145	7	0.9	505	9	US-09-604-287A-478	Sequence 478, App
146	7	0.9	505	9	US-09-604-287A-485	Sequence 485, App
147	7	0.9	505	11	US-09-551-621-478	Sequence 478, App
148	7	0.9	505	12	US-10-124-805-478	Sequence 478, App
149	7	0.9	505	12	US-10-124-805-485	Sequence 485, App
150	7	0.9	505	14	US-10-007-805-478	Sequence 478, App
151	7	0.9	505	14	US-10-007-805-485	Sequence 485, App
152	7	0.9	505	15	US-10-076-622-478	Sequence 478, App
153	7	0.9	511	15	US-10-076-622-485	Sequence 485, App
154	7	0.9	511	15	US-09-796-256A-4	Sequence 4, App
155	7	0.9	511	15	US-10-128-714-8150	Sequence 8150, App
156	7	0.9	511	15	US-10-128-714-8150	Sequence 8150, App
157	7	0.9	512	9	US-09-977-260-16	Sequence 16, App
158	7	0.9	512	10	US-09-977-260-16	Sequence 16, App
159	7	0.9	512	11	US-09-977-261-16	Sequence 16, App
160	7	0.9	518	15	US-10-313-852-1	Sequence 1, App
161	7	0.9	518	15	US-10-313-852-3	Sequence 3, App

162	7	0.9	518	15	US-10-314-033-1	Sequence 1, App
163	7	0.9	518	15	US-10-314-033-3	Sequence 3, App
164	7	0.9	529	16	US-10-174-693-405	Sequence 405, App
165	7	0.9	531	15	US-10-156-761-12621	Sequence 12621, A
166	7	0.9	534	15	US-09-801-368-124	Sequence 124, App
167	7	0.9	538	15	US-10-043-487-329	Sequence 329, App
168	7	0.9	582	15	US-10-156-761-11938	Sequence 11938, A
169	7	0.9	593	15	US-10-156-761-11938	Sequence 11938, A
170	7	0.9	594	15	US-10-106-698-6185	Sequence 6185, App
171	7	0.9	631	15	US-10-128-714-3111	Sequence 3111, App
172	7	0.9	659	12	US-10-363-427-12	Sequence 12, App
173	7	0.9	660	14	US-10-067-385-6	Sequence 6, App
174	7	0.9	670	15	US-10-156-761-14107	Sequence 14107, A
175	7	0.9	677	9	US-09-815-242-11991	Sequence 11991, A
176	7	0.9	679	10	US-09-738-626-5995	Sequence 5995, App
177	7	0.9	694	11	US-09-842-758-75	Sequence 75, App
178	7	0.9	694	12	US-10-345-680-26	Sequence 26, App
179	7	0.9	720	12	US-10-363-427-8	Sequence 8, App
180	7	0.9	740	9	US-09-815-242-10376	Sequence 10376, A
181	7	0.9	754	15	US-10-128-714-3448	Sequence 3448, App
182	7	0.9	775	9	US-09-815-242-13811	Sequence 13811, A
183	7	0.9	779	9	US-09-815-242-12093	Sequence 12093, A
184	7	0.9	788	12	US-10-032-585-7007	Sequence 7007, App
185	7	0.9	792	15	US-10-128-714-8111	Sequence 8111, App
186	7	0.9	799	15	US-10-156-761-9985	Sequence 9985, App
187	7	0.9	819	10	US-09-882-529-3	Sequence 2, App
188	7	0.9	821	10	US-09-882-529-3	Sequence 2, App
189	7	0.9	831	9	US-09-815-242-10520	Sequence 10520, A
190	7	0.9	856	10	US-09-882-529-4	Sequence 4, App
191	7	0.9	870	9	US-09-815-242-4993	Sequence 5493, App
192	7	0.9	870	9	US-09-815-242-12637	Sequence 12637, A
193	7	0.9	879	9	US-09-872-733-16	Sequence 16, App
194	7	0.9	879	15	US-10-263-020-16	Sequence 16, App
195	7	0.9	886	8	US-08-781-986A-5235	Sequence 535, App
196	7	0.9	888	11	US-09-884-465A-338	Sequence 338, App
197	7	0.9	889	11	US-09-884-465A-338	Sequence 338, App
198	7	0.9	894	11	US-09-884-465A-337	Sequence 337, App
199	7	0.9	894	11	US-09-884-465A-340	Sequence 340, App
200	7	0.9	894	11	US-09-884-465A-342	Sequence 342, App

RESULT 1
US-09-765-272-56
Sequence 56, Application US/09765272
Patent No. US20020061545A1
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/765, 272
FILING DATE: 22-Jan-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 796 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 56:
US-09-765-272-56

Query Match 99.9%; Score 795; DB 9; Length 796;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 SYELGLYQARTVKNRVSYIDGKATQKTENTLPPEVSKREGINAEQIVIKITTDGQYVT 60
1 SYELGLYQARTVKNRVSYIDGKATQKTENTLPPEVSKREGINAEQIVIKITTDGQYVT 60
61 SHGDHYHYNGKVPYDAIISELLMKDPNYKLMKEDIIVNEVKGYIKNVGGKYYVYLKDA 120
61 SHGDHYHYNGKVPYDAIISELLMKDPNYKLMKEDIIVNEVKGYIKNVGGKYYVYLKDA 120
121 AHADNRTKEEINROKQEHQREHGTFRNDGVALARSGRTTDDGYIFNADIIEDT 180
121 AHADNRTKEEINROKQEHQREHGTFRNDGVALARSGRTTDDGYIFNADIIEDT 180
181 GDAYIVPHGDHYHYPINELASASELAEEAFLSGRGLNSRTYRRONSNTSRTNWPVS 240
181 GDAYIVPHGDHYHYPINELASASELAEEAFLSGRGLNSRTYRRONSNTSRTNWPVS 240
241 VSNPGTNTNTSNNSNTNSQASQNDISLLKQLYKLPLORHVESDGLVFDEPAQITSR 300
241 VSNPGTNTNTSNNSNTNSQASQNDISLLKQLYKLPLORHVESDGLVFDEPAQITSR 300
301 ARGVAVPHGDHYHYPISQMSLEERLARIIPRYSNHWPDSRPEQSPQTPPEPSPG 360
301 ARGVAVPHGDHYHYPISQMSLEERLARIIPRYSNHWPDSRPEQSPQTPPEPSPG 360
361 POPAPNLIKIDSNSSLSQLVKRVKGEVPEEKISRVFAKDPSEFVKNLESKLSKQES 420
361 POPAPNLIKIDSNSSLSQLVKRVKGEVPEEKISRVFAKDPSEFVKNLESKLSKQES 420
421 VSHTLTAKKENVAPRQDEFYDKAYNLLTEAHKALFNKKGNSDPQALDKLERLNDESTN 480
421 VSHTLTAKKENVAPRQDEFYDKAYNLLTEAHKALFNKKGNSDPQALDKLERLNDESTN 480
481 KEKLVNDLLAFAPITHEPERLGRPNQOIEYTEDEVRIQAOLADYTTSDGYIFPEHDIISD 540
481 KEKLVNDLLAFAPITHEPERLGRPNQOIEYTEDEVRIQAOLADYTTSDGYIFPEHDIISD 540
541 EGDAYVTPHMGSHMWGKDSLSPKEKVAQAQYTKESKILPSSPDADVKAPPTDSSAAIY 600
541 EGDAYVTPHMGSHMWGKDSLSPKEKVAQAQYTKESKILPSSPDADVKAPPTDSSAAIY 600
601 NRKVGSKRIPLVRLPYMVEHTEVKNGNLLIPKDHYNIKFAMFDDHTYKANGYTLSD 660
601 NRKVGSKRIPLVRLPYMVEHTEVKNGNLLIPKDHYNIKFAMFDDHTYKANGYTLSD 660
661 LEFTIYYVHEHPERPHSNDGMGNASEHVLGKKDHSDEPNKFKADEEPEVEETPADEVP 720
661 LEFTIYYVHEHPERPHSNDGMGNASEHVLGKKDHSDEPNKFKADEEPEVEETPADEVP 720
721 QVTEKEVAQKAEAVLLAKVTDSLKANATELLAGRNLLTIOIDMNNSIMAEAEKLLA 780
721 QVTEKEVAQKAEAVLLAKVTDSLKANATELLAGRNLLTIOIDMNNSIMAEAEKLLA 780
781 LKGSNPSVSKKIN 796
|||||

Db 781 LKGSNPSVSKKIN 796

RESULT 2
US-09-884-465A-7
Sequence 7; Application US/09884465A
Publication No. US20030077293A1
GENERAL INFORMATION:
APPLICANT: Shire Biochem, Inc.
APPLICANT: Hamel, Josee
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
PRIOR FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 7
LENGTH: 840
TYPE: PR
ORGANISM: Streptococcus pneumoniae
US-09-884-465A-7

Query Match 10.1%; Score 80; DB 11; Length 840;
Best Local Similarity 100.0%; Pred. No. 6e-70;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

11 TVKENNRVSYIDGKATQKTENTLPPEVSKREGINAEQIVIKITTDGQYVTSQDHYHYN 70
11 TVKENNRVSYIDGKATQKTENTLPPEVSKREGINAEQIVIKITTDGQYVTSQDHYHYN 70
31 TVKENNRVSYIDGKATQKTENTLPPEVSKREGINAEQIVIKITTDGQYVTSQDHYHYN 90
31 TVKENNRVSYIDGKATQKTENTLPPEVSKREGINAEQIVIKITTDGQYVTSQDHYHYN 90
71 GKVPYDAIISELLMKDPNY 90
71 GKVPYDAIISELLMKDPNY 90
91 GKVPYDAIISELLMKDPNY 110
91 GKVPYDAIISELLMKDPNY 110

RESULT 3
US-09-769-787-194
Sequence 194; Application US/09769787
Publication No. US20030091577A1
GENERAL INFORMATION:
APPLICANT: Microbial Technics Limited
APPLICANT: Gilbert, Christophe FG
APPLICANT: Hansbro, Philip M
TITLE OF INVENTION: Proteins
FILE REFERENCE: PWC/P21129WO
CURRENT APPLICATION NUMBER: US/09/769,787
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: GB 9816337.1
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: US 60/125164
PRIOR FILING DATE: 1999-03-19
NUMBER OF SEQ ID NOS: 388
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 194
LENGTH: 826
TYPE: PR
ORGANISM: Streptococcus pneumoniae
US-09-769-787-194

Query Match 7.5%; Score 60; DB 11; Length 826;
Best Local Similarity 100.0%; Pred. No. 3.6e-50;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

31 ENLTPDEVSKREGINAEQIVIKITTDGQYVTSQDHYHYNGKVPYDAIISELLMKDPNY 90
31 ENLTPDEVSKREGINAEQIVIKITTDGQYVTSQDHYHYNGKVPYDAIISELLMKDPNY 90
52 ENLTPDEVSKREGINAEQIVIKITTDGQYVTSQDHYHYNGKVPYDAIISELLMKDPNY 111
52 ENLTPDEVSKREGINAEQIVIKITTDGQYVTSQDHYHYNGKVPYDAIISELLMKDPNY 111

RESULT 4
US-09-884-465A-8
Sequence 8, Application US/09884465A
Publication No. US2003007293A1
GENERAL INFORMATION:
APPLICANT: Shire Biochem, Inc.
APPLICANT: Hamel, Josee
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 8
LENGTH: 838
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-884-465A-8

Query Match
Best Local Similarity 100.0%; Score 60; DB 11; Length 838;
Best Local Similarity 100.0%; Pred. No. 3,6e-50;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 31 ENLTDEVSKREGINAEOIVKITDQGYVTSQDHYHYNGKVPYDAIISBELMKDPNY 90
Db 52 ENLTDEVSKREGINAEOIVKITDQGYVTSQDHYHYNGKVPYDAIISBELMKDPNY 111

RESULT 5
US-09-765-272-66
Sequence 66, Application US/09765272
Patent No. US20020061545A1
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/765,272
FILING DATE: 22-Jan-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 763 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 66:
US-09-765-272-66

Query Match
Best Local Similarity 100.0%; Score 51; DB 9; Length 763;
Best Local Similarity 100.0%; Pred. No. 2,7e-41;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 160 QGRYTTDDGYIFNNSDIETDGDAYIVPHGHHYIPNNELASBELAAEA 210
Db 159 QGRYTTDDGYIFNNSDIETDGDAYIVPHGHHYIPNNELASBELAAEA 209

RESULT 6
US-09-884-465A-12
Sequence 12, Application US/09884465A
Publication No. US2003007293A1
GENERAL INFORMATION:
APPLICANT: Shire Biochem, Inc.
APPLICANT: Hamel, Josee
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 12
LENGTH: 42
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism
US-09-884-465A-12

Query Match
Best Local Similarity 100.0%; Score 32; DB 11; Length 42;
Best Local Similarity 100.0%; Pred. No. 1,1e-23;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 46 AEQIVIKITDQGYVTSQDHYHYNGKVPYDA 77
Db 4 AEQIVIKITDQGYVTSQDHYHYNGKVPYDA 35

RESULT 7
US-09-765-272-182
Sequence 182, Application US/09765272
Patent No. US20020061545A1
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/765,272

```

; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 182:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 447 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 182:
US-09-765-272-182
```

```

Query Match          4.0%; Score 32; DB 9; Length 447;
Best Local Similarity 100.0%; Pred. No. 1e-22;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 46 AEQIVIKITDQGYVTSHGHDHYHYNGKVPYDA 77
DB 43 AEQIVIKITDQGYVTSHGHDHYHYNGKVPYDA 74
```

```

RESULT 8
US-09-769-787-38
; Sequence 38, Application US/09769787
; Publication No. US20030091577A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PMC/P21128WO
; CURRENT APPLICATION NUMBER: US/09/769,787
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 38
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-769-787-38
```

```

Query Match          4.0%; Score 32; DB 11; Length 484;
Best Local Similarity 100.0%; Pred. No. 1.1e-22;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 46 AEQIVIKITDQGYVTSHGHDHYHYNGKVPYDA 77
DB 67 AEQIVIKITDQGYVTSHGHDHYHYNGKVPYDA 98
```

```

RESULT 9
US-09-769-744A-24
; Sequence 24, Application US/09769744A
; Publication No. US20030134407A1
; GENERAL INFORMATION:
; APPLICANT: Ie Page, Richard WF
; APPLICANT: Wells, Jeremy M
; APPLICANT: Hamblift, Sean B
; APPLICANT: Hansbro, Philip M
```

```

; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PMC/P21122WO
; CURRENT APPLICATION NUMBER: US/09/769,744A
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: PCT/GB99/02452
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: GB 9816336.3
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125329
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-769-744A-24
```

```

Query Match          4.0%; Score 32; DB 12; Length 484;
Best Local Similarity 100.0%; Pred. No. 1.1e-22;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 46 AEQIVIKITDQGYVTSHGHDHYHYNGKVPYDA 77
DB 67 AEQIVIKITDQGYVTSHGHDHYHYNGKVPYDA 98
```

```

RESULT 10
US-09-884-465A-10
; Sequence 10, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Joseph
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 840
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-10
```

```

Query Match          4.0%; Score 32; DB 11; Length 840;
Best Local Similarity 100.0%; Pred. No. 1.8e-22;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 46 AEQIVIKITDQGYVTSHGHDHYHYNGKVPYDA 77
DB 47 AEQIVIKITDQGYVTSHGHDHYHYNGKVPYDA 78
```

```

RESULT 11
US-09-884-465A-6
; Sequence 6, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Joseph
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
```

```

; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 1039
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; US-09-884-465A-6

```

```

Query Match          4.0%; Score 32; DB 11; Length 1039;
Best Local Similarity 100.0%; Pred. No. 2,2e-22;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy      46 AEOIVIKITDQGYVTSKGDHYYHYNGKVPYDA 77
      |||||
Db      67 AEOIVIKITDQGYVTSKGDHYYHYNGKVPYDA 98

```

```

RESULT 12
; US-09-884-465A-380
; Sequence 380, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 380
; LENGTH: 1139
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: Xaa = Methionine or nothing
; NAME/KEY: MISC_FEATURE
; LOCATION: (570)..(570)
; OTHER INFORMATION: Xaa = Glycine or nothing
; NAME/KEY: MISC_FEATURE
; LOCATION: (571)..(571)
; OTHER INFORMATION: Xaa = Proline or nothing
; US-09-884-465A-380

```

```

Query Match          3.4%; Score 27; DB 11; Length 1139;
Best Local Similarity 100.0%; Pred. No. 2,1e-17;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy      326 RIARIIPLRYSNMHWVPSRPEQSPQ 352
      |||||
Db      63 RIARIIPLRYSNMHWVPSRPEQSPQ 89

```

```

RESULT 13
; US-09-884-465A-378
; Sequence 378, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.

```

```

; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 378
; LENGTH: 1378
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: Xaa = Methionine or nothing
; NAME/KEY: MISC_FEATURE
; LOCATION: (570)..(570)
; OTHER INFORMATION: Xaa = Glycine or nothing
; NAME/KEY: MISC_FEATURE
; LOCATION: (571)..(571)
; OTHER INFORMATION: Xaa = Proline or nothing
; US-09-884-465A-378

```

```

Query Match          3.4%; Score 27; DB 11; Length 1378;
Best Local Similarity 100.0%; Pred. No. 2,6e-17;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy      326 RIARIIPLRYSNMHWVPSRPEQSPQ 352
      |||||
Db      63 RIARIIPLRYSNMHWVPSRPEQSPQ 89

```

```

RESULT 14
; US-09-884-465A-376
; Sequence 376, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 376
; LENGTH: 999
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: Xaa = Methionine or nothing
; NAME/KEY: MISC_FEATURE
; LOCATION: (430)..(430)
; OTHER INFORMATION: Xaa = Glycine or nothing
; NAME/KEY: MISC_FEATURE
; LOCATION: (431)..(431)
; OTHER INFORMATION: Xaa = Proline or nothing

```

US-09-884-465A-376

Query Match 3.0%; Score 24; DB 11; Length 999;

Best Local Similarity 100.0%; Pred. No. 1.8e-14;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 266 DIDSLLKQLYKPLPSQRHVESDGL 289

DB 2 DIDSLLKQLYKPLPSQRHVESDGL 25

RESULT 15

US-09-884-465A-377

Sequence 377, Application US/09884465A

Publication No. US20030077293A1

GENERAL INFORMATION:

APPLICANT: Shire Biochem, Inc.

APPLICANT: Hamel, Josee

APPLICANT: Brodeur, Bernard

APPLICANT: Martin, Denis

APPLICANT: Charland, Nathalie

APPLICANT: Ouellet, Catherine

TITLE OF INVENTION: Streptococcus Antigens

FILE REFERENCE: 055190-0044

CURRENT APPLICATION NUMBER: US/09/884,465A

PRIOR FILING DATE: 2001-06-20

PRIOR APPLICATION NUMBER: 60/212,683

PRIOR FILING DATE: 2000-06-20

NUMBER OF SEQ ID NOS: 384

SOFTWARE: PatentIn version 3.1

SEQ ID NO 377

LENGTH: 999

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Unknown Organism

NAME/KEY: MISC_FEATURE

LOCATION: (1)..(1)

OTHER INFORMATION: Xaa = Methionine or nothing

NAME/KEY: MISC_FEATURE

LOCATION: (570)..(570)

OTHER INFORMATION: Xaa = Glycine or nothing

NAME/KEY: MISC_FEATURE

LOCATION: (571)..(571)

OTHER INFORMATION: Xaa = Proline or nothing

US-09-884-465A-377

Query Match 3.0%; Score 24; DB 11; Length 999;
Best Local Similarity 100.0%; Pred. No. 1.8e-14;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 266 DIDSLLKQLYKPLPSQRHVESDGL 289

DB 572 DIDSLLKQLYKPLPSQRHVESDGL 595

RESULT 16

US-09-884-465A-383

Sequence 383, Application US/09884465A

Publication No. US20030077293A1

GENERAL INFORMATION:

APPLICANT: Shire Biochem, Inc.

APPLICANT: Hamel, Josee

APPLICANT: Brodeur, Bernard

APPLICANT: Martin, Denis

APPLICANT: Charland, Nathalie

APPLICANT: Ouellet, Catherine

TITLE OF INVENTION: Streptococcus Antigens

FILE REFERENCE: 055190-0044

CURRENT APPLICATION NUMBER: US/09/884,465A

PRIOR FILING DATE: 2001-06-20

PRIOR APPLICATION NUMBER: 60/212,683

NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 383
LENGTH: 1126
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism
NAME/KEY: MISC_FEATURE
LOCATION: (1)..(1)
OTHER INFORMATION: Xaa = Methionine or nothing
NAME/KEY: MISC_FEATURE
LOCATION: (557)..(557)
OTHER INFORMATION: Xaa = Glycine or nothing
NAME/KEY: MISC_FEATURE
LOCATION: (558)..(558)
OTHER INFORMATION: Xaa = Proline or nothing
US-09-884-465A-383Query Match 3.0%; Score 24; DB 11; Length 1126;
Best Local Similarity 100.0%; Pred. No. 2e-14;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 266 DIDSLLKQLYKPLPSQRHVESDGL 289

DB 2 DIDSLLKQLYKPLPSQRHVESDGL 25

RESULT 17

US-09-884-465A-381

Sequence 381, Application US/09884465A

Publication No. US20030077293A1

GENERAL INFORMATION:

APPLICANT: Shire Biochem, Inc.

APPLICANT: Hamel, Josee

APPLICANT: Brodeur, Bernard

APPLICANT: Martin, Denis

APPLICANT: Charland, Nathalie

APPLICANT: Ouellet, Catherine

TITLE OF INVENTION: Streptococcus Antigens

FILE REFERENCE: 055190-0044

CURRENT APPLICATION NUMBER: US/09/884,465A

PRIOR FILING DATE: 2001-06-20

PRIOR APPLICATION NUMBER: 60/212,683

PRIOR FILING DATE: 2000-06-20

NUMBER OF SEQ ID NOS: 384

SOFTWARE: PatentIn version 3.1

SEQ ID NO 381

LENGTH: 1238

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Unknown Organism

NAME/KEY: MISC_FEATURE

LOCATION: (1)..(1)

OTHER INFORMATION: Xaa = Methionine or nothing

NAME/KEY: MISC_FEATURE

LOCATION: (430)..(430)

OTHER INFORMATION: Xaa = Glycine or nothing

NAME/KEY: MISC_FEATURE

LOCATION: (431)..(431)

OTHER INFORMATION: Xaa = Proline or nothing

US-09-884-465A-381

Query Match 3.0%; Score 24; DB 11; Length 1238;
Best Local Similarity 100.0%; Pred. No. 2.2e-14;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 266 DIDSLLKQLYKPLPSQRHVESDGL 289

DB 2 DIDSLLKQLYKPLPSQRHVESDGL 25

RESULT 18

US-09-884-465A-382
; Sequence 382, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 382
; LENGTH: 1365
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
; NAME/KEY: MISC FEATURE
; LOCATION: (1)-(1)
; OTHER INFORMATION: Xaa = Methionine or nothing
; NAME/KEY: MISC FEATURE
; LOCATION: (557)-(557)
; OTHER INFORMATION: Xaa = Glycine or nothing
; NAME/KEY: MISC FEATURE
; LOCATION: (558)-(558)
; OTHER INFORMATION: Xaa = Proline or nothing
US-09-884-465A-382

Query Match

Best Local Similarity 3.0%; Score 24; DB 11; Length 1365;
Best Local Similarity 100.0%; Pred. No. 2,4e-14;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 266 DIDSLLKQLYKLPISQRHVESDGL 289
|||
Db 2 DIDSLLKQLYKLPISQRHVESDGL 25

RESULT 19

US-09-884-465A-11
; Sequence 11, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 94
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-11

Query Match 2.9%; Score 23; DB 11; Length 94;

Best Local Similarity 100.0%; Pred. No. 1.9e-14;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 DQGVTSQHDHYHYNGKVPYDA 77
|||
Db 1 DQGVTSQHDHYHYNGKVPYDA 23

RESULT 20

US-09-769-736-24
; Sequence 24, Application US/09769736
; Publication No. US20030138775A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Le Page, Richard WF
; APPLICANT: Wells, Jeremy M
; APPLICANT: Hamilty, Sean B
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PMC/P21089wo
; CURRENT APPLICATION NUMBER: US/09/769,736
; PRIOR FILING DATE: 2003-02-14
; PRIOR APPLICATION NUMBER: GB 9816335.5
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125163
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 381
; TYPE: PRT
; ORGANISM: Streptococcus agalactiae
US-09-769-736-24

Query Match 2.1%; Score 17; DB 12; Length 381;
Best Local Similarity 100.0%; Pred. No. 6e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 69 YNGKVPYDAIISELLM 85
|||
Db 92 YNGKVPYDAIISELLM 108

RESULT 21

US-09-252-088-15
; Sequence 15, Application US/09252088
; Publication No. US20030031682A1
; GENERAL INFORMATION:
; APPLICANT: BRODEUR, Bernard R.
; APPLICANT: RIOUX, Clement
; APPLICANT: BOYER, Martine
; APPLICANT: CHARLEBOIS, Isabelle
; APPLICANT: HAMEL, Josee
; APPLICANT: MARTIN, Denis
; TITLE OF INVENTION: NOVEL GROUP B STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 8331-9002
; CURRENT APPLICATION NUMBER: US/09/252,088
; PRIOR FILING DATE: 1998-02-18
; EARLIER APPLICATION NUMBER: US/60/075,425
; PRIOR FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 793
; TYPE: PRT
; ORGANISM: group B streptococcus
US-09-252-088-15

Query Match 2.1%; Score 17; DB 11; Length 793;
Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 69 YNGKVPYDAIISELLM 85
|||

Db 63 YNGKVPYDAIISELLM 79

RESULT 22

US-09-769-736-18
; Sequence 18, Application US/09769736
; Publication No. US20030138775A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Le Page, Richard WF
; APPLICANT: Wells, Jeremy M
; APPLICANT: Hannify, Sean B
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PMC/P21089WO
; CURRENT APPLICATION NUMBER: US/09/769,736
; CURRENT FILING DATE: 2003-02-14
; PRIOR APPLICATION NUMBER: GB 9816335.5
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125163
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 18
; LENGTH: 822
; TYPE: PRT
; ORGANISM: Streptococcus agalactiae
US-09-769-736-18

Query Match 2.1%; Score 17; DB 12; Length 822;

Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 69 YNGKVPYDAIISELLM 85

Db 92 YNGKVPYDAIISELLM 108

RESULT 23

US-09-252-088-16
; Sequence 16, Application US/09252088
; Publication No. US20030031682A1
; GENERAL INFORMATION:
; APPLICANT: BRODEUR, Bernard R.
; APPLICANT: RIOUX, Clement
; APPLICANT: BOYER, Martine
; APPLICANT: CHARLEBOIS, Isabelle
; APPLICANT: HAMEL, Jose
; TITLE OF INVENTION: NOVEL GROUP B STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 8331-9002
; CURRENT APPLICATION NUMBER: US/09/252,088
; CURRENT FILING DATE: 1999-02-18
; EARLIER APPLICATION NUMBER: US/60/075,425
; EARLIER FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 16
; LENGTH: 715
; TYPE: PRT
; ORGANISM: Group B streptococcus
US-09-252-088-16

Query Match 1.5%; Score 12; DB 11; Length 715;

Best Local Similarity 100.0%; Pred. No. 0.0096;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 160 QGRYTTDDGYIF 171

Db 84 QGRYTTDDGYIF 95

RESULT 24

US-09-884-465A-384

; Sequence 384, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 384
; LENGTH: 913
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: Xaa = Methionine or nothing
; NAME/KEY: MISC_FEATURE
; LOCATION: (345)..(345)
; OTHER INFORMATION: Xaa = Proline or nothing
US-09-884-465A-384

Query Match 1.5%; Score 12; DB 11; Length 913;

Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 539 SDEGDVYVTPHM 550

Db 54 SDEGDVYVTPHM 65

RESULT 25

US-09-884-465A-379
; Sequence 379, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 379
; LENGTH: 1152
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: Xaa = Methionine or nothing
; NAME/KEY: MISC_FEATURE
; LOCATION: (344)..(344)

OTHER INFORMATION: Xaa = Glycine or nothing
NAME/KEY: MISC FEATURE
LOCATION: (345)...(345)
OTHER INFORMATION: Xaa = Proline or nothing
US-09-884-465A-379

Query Match 1.5%; Score 12; DB 11; Length 1152;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 307 PHGDHYHPIPS 318
DB 421 PHGDHYHPIPS 432

RESULT 26
US-09-769-736-72
Sequence 72, Application US/09769736
Publication No. US2003013875A1
GENERAL INFORMATION:
APPLICANT: Microbial Technics Limited
APPLICANT: Le Page, Richard WF
APPLICANT: Wells, Jeremy M
APPLICANT: Hamifly, Sean B
TITLE OF INVENTION: Proteins
FILE REFERENCE: PMC/P21089wo
CURRENT APPLICATION NUMBER: US/09/769,736
CURRENT FILING DATE: 2003-02-14
PRIOR APPLICATION NUMBER: GB 9816335.5
PRIOR FILING DATE: 1998-07-27
PRIOR APPLICATION NUMBER: US 60/125163
PRIOR FILING DATE: 1999-03-19
NUMBER OF SEQ ID NOS: 212
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 72
LENGTH: 485
TYPE: PRT
ORGANISM: Streptococcus agalactiae
US-09-769-736-72

Query Match 1.4%; Score 11; DB 12; Length 485;
Best Local Similarity 100.0%; Pred. No. 0.065;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 523 KYTSDGYIFD 533
DB 110 KYTSDGYIFD 120

RESULT 27
US-09-884-465A-300
Sequence 300, Application US/09884465A
Publication No. US20030077293A1
GENERAL INFORMATION:
APPLICANT: Shire Biochem, Inc.
APPLICANT: Hamel, Josee
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 300
LENGTH: 272
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:

OTHER INFORMATION: Unknown Organism
US-09-884-465A-300

Query Match 1.1%; Score 9; DB 11; Length 272;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 634 KDHYNIKF 642
DB 110 KDHYNIKF 118

RESULT 28
US-09-884-465A-22
Sequence 22, Application US/09884465A
Publication No. US20030077293A1
GENERAL INFORMATION:
APPLICANT: Shire Biochem, Inc.
APPLICANT: Hamel, Josee
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 22
LENGTH: 86
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism
US-09-884-465A-22

Query Match 1.0%; Score 8; DB 11; Length 86;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 635 DHYNIKF 642
DB 52 DHYNIKF 59

RESULT 29
US-09-884-465A-374
Sequence 374, Application US/09884465A
Publication No. US20030077293A1
GENERAL INFORMATION:
APPLICANT: Shire Biochem, Inc.
APPLICANT: Hamel, Josee
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 374
LENGTH: 259
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism

US-09-884-465A-374

Query Match 1.0%; Score 8; DB 11; Length 259;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 635 DHYHNIKF 642
|||
Db 98 DHYHNIKF 105

RESULT 30

US-09-884-465A-302
; Sequence 302, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 302
; LENGTH: 270
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-302

Query Match 1.0%; Score 8; DB 11; Length 270;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 635 DHYHNIKF 642
|||
Db 109 DHYHNIKF 116

RESULT 31

US-09-884-465A-258
; Sequence 258, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 258
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-258

Query Match 1.0%; Score 8; DB 11; Length 272;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 635 DHYHNIKF 642
|||
Db 111 DHYHNIKF 118

RESULT 32

US-09-884-465A-293
; Sequence 293, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 293
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-293

Query Match 1.0%; Score 8; DB 11; Length 272;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 635 DHYHNIKF 642
|||
Db 111 DHYHNIKF 118

RESULT 33

US-10-156-761-13480
; Sequence 13480, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13480
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-13480

Query Match 1.0%; Score 8; DB 15; Length 400;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 669 VEHDPERP 676
 |||||
 Db 70 VEHDPERP 77

RESULT 34

US-09-815-242-5401
 ; Sequence 5401, Application US/09815242
 ; Patent No. US2002061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 5401
 ; LENGTH: 482
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus aureus
 ; US-09-815-242-5401

Query Match 1.0%; Score 8; DB 9; Length 482;
 Best Local Similarity 100.0%; Pred. No. 60;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 150 NDGAVALA 157
 |||||
 Db 69 NDGAVALA 76

RESULT 35

US-09-815-242-12564
 ; Sequence 12564, Application US/09815242
 ; Patent No. US2002061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: Prokaryotes

FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 12564
 ; LENGTH: 488
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus aureus
 ; US-09-815-242-12564

Query Match 1.0%; Score 8; DB 9; Length 488;
 Best Local Similarity 100.0%; Pred. No. 61;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 150 NDGAVALA 157
 |||||
 Db 72 NDGAVALA 79

RESULT 36

US-09-884-465A-336
 ; Sequence 336, Application US/09884465A
 ; Publication No. US20030077293A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Shire Biochem, Inc.
 ; APPLICANT: Hamel, Josee
 ; APPLICANT: Brodeur, Bernard
 ; APPLICANT: Martin, Denis
 ; APPLICANT: Charland, Nathalie
 ; APPLICANT: Ouellet, Catherine
 ; TITLE OF INVENTION: Streptococcus Antigenic
 ; FILE REFERENCE: 055190-0044
 ; CURRENT APPLICATION NUMBER: US/09/884,465A
 ; CURRENT FILING DATE: 2001-06-20
 ; PRIOR APPLICATION NUMBER: 60/212,683
 ; PRIOR FILING DATE: 2000-06-20
 ; NUMBER OF SEQ ID NOS: 384
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 336
 ; LENGTH: 894
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Unknown Organism
 ; US-09-884-465A-336

Query Match 1.0%; Score 8; DB 11; Length 894;
 Best Local Similarity 100.0%; Pred. No. 11e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 635 DRYHNKIF 642
 |||||
 Db 111 DRYHNKIF 118

RESULT 37

US-09-884-465A-344
 ; Sequence 344, Application US/09884465A

```
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 344
; LENGTH: 895
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-344
```

```
Query Match          1.0%; Score 8; DB 11; Length 895;
Best Local Similarity 100.0%; Pred.No. 1.1e+02;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;
```

```
QY      635 DHHYNIKF 642
        |||||||
Db       111 DHHYNIKF 118
```

```
RESULT 38
US-09-884-465A-335
; Sequence 335, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 335
; LENGTH: 900
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-335
```

```
Query Match          1.0%; Score 8; DB 11; Length 900;
Best Local Similarity 100.0%; Pred.No. 1.1e+02;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;
```

```
QY      635 DHHYNIKF 642
        |||||||
Db       111 DHHYNIKF 118
```

```
RESULT 39
US-09-884-465A-343
; Sequence 343, Application US/09884465A
; Publication No. US20030077293A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 343
; LENGTH: 901
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-343
```

```
Query Match          1.0%; Score 8; DB 11; Length 901;
Best Local Similarity 100.0%; Pred.No. 1.1e+02;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;
```

```
QY      635 DHHYNIKF 642
        |||||||
Db       111 DHHYNIKF 118
```

```
RESULT 40
US-09-884-465A-332
; Sequence 332, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 332
; LENGTH: 906
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-332
```

```
Query Match          1.0%; Score 8; DB 11; Length 906;
Best Local Similarity 100.0%; Pred.No. 1.1e+02;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;
```

```
QY      635 DHHYNIKF 642
        |||||||
Db       111 DHHYNIKF 118
```

```
RESULT 41
US-09-884-465A-369
; Sequence 369, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
```

APPLICANT: Shire Biochem, Inc.
APPLICANT: Hamel, Josee
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 369
LENGTH: 906
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism
US-09-884-465A-369

Query Match 1.0%; Score 8; DB 11; Length 906;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 635 DHYNIKF 642
DB 745 DHYNIKF 752

RESULT 42
US-09-884-465A-370
Sequence 370, Application US/09884465A
Publication No. US20030077293A1
GENERAL INFORMATION:
APPLICANT: Shire Biochem, Inc.
APPLICANT: Hamel, Josee
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 370
LENGTH: 906
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism
US-09-884-465A-370

Query Match 1.0%; Score 8; DB 11; Length 906;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 635 DHYNIKF 642
DB 111 DHYNIKF 118

RESULT 43
US-09-884-465A-371
Sequence 371, Application US/09884465A
Publication No. US20030077293A1
GENERAL INFORMATION:
APPLICANT: Shire Biochem, Inc.

APPLICANT: Hamel, Josee
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 371
LENGTH: 906
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism
US-09-884-465A-371

Query Match 1.0%; Score 8; DB 11; Length 906;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 635 DHYNIKF 642
DB 745 DHYNIKF 752

RESULT 44
US-09-884-465A-372
Sequence 372, Application US/09884465A
Publication No. US20030077293A1
GENERAL INFORMATION:
APPLICANT: Shire Biochem, Inc.
APPLICANT: Hamel, Josee
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 372
LENGTH: 906
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism
US-09-884-465A-372

Query Match 1.0%; Score 8; DB 11; Length 906;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 635 DHYNIKF 642
DB 111 DHYNIKF 118

RESULT 45
US-09-884-465A-373
Sequence 373, Application US/09884465A
Publication No. US20030077293A1
GENERAL INFORMATION:
APPLICANT: Shire Biochem, Inc.
APPLICANT: Hamel, Josee

APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
PRIOR FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 373
LENGTH: 906
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism
US-09-884-465A-373

Query Match 1.0%; Score 8; DB 11; Length 906;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 635 DHHYNIKF 642
|||||
DB 745 DHHYNIKF 752

RESULT 46
US-10-128-714-3262
Sequence 3262, Application US/10128714
Publication No. US20030119013A1
GENERAL INFORMATION:
APPLICANT: Jiang, Bo
APPLICANT: Hu, Wengji
APPLICANT: Tishkoff, Daniel
APPLICANT: Zamudio, Carlos
APPLICANT: Eroshkin, Alexey M
APPLICANT: Lemieux, Sebastien M
TITLE OF INVENTION: Identification of Essential Genes in *Aspergillus fumigatus* and
TITLE OF INVENTION: Methods of Use
FILE REFERENCE: 10182-018-999
CURRENT APPLICATION NUMBER: US/10/128,714
CURRENT FILING DATE: 2002-04-23
PRIOR APPLICATION NUMBER: US 60/285,697
PRIOR FILING DATE: 2001-04-23
PRIOR APPLICATION NUMBER: US 60/287,066
PRIOR FILING DATE: 2001-04-27
PRIOR APPLICATION NUMBER: US 60/295,890
PRIOR FILING DATE: 2001-06-05
PRIOR APPLICATION NUMBER: US 60/303,899
PRIOR FILING DATE: 2001-07-09
PRIOR APPLICATION NUMBER: US 60/316,362
NUMBER OF SEQ ID NOS: 8603
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3262
LENGTH: 1104
TYPE: PRT
ORGANISM: *Aspergillus fumigatus*
US-10-128-714-3262

Query Match 1.0%; Score 8; DB 15; Length 1104;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 707 EEPVEETP 714
|||||
DB 952 EEPVEETP 959

RESULT 47

US-09-712-363-259
Sequence 259, Application US/09712363
Patent No. US20020164588A1
GENERAL INFORMATION:
APPLICANT: Eisenberg, David
APPLICANT: Rotstein, Sergio H.
APPLICANT: Marcotte, Edward M.
TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
FILE REFERENCE: 07419-032001
CURRENT APPLICATION NUMBER: US/09/712,363
CURRENT FILING DATE: 2000-11-13
PRIOR APPLICATION NUMBER: PCT/US00/02246
PRIOR FILING DATE: 2000-01-28
PRIOR APPLICATION NUMBER: 60/179,531
PRIOR FILING DATE: 2000-02-01
PRIOR APPLICATION NUMBER: 60/117,844
PRIOR FILING DATE: 1999-01-29
PRIOR APPLICATION NUMBER: 60/118,206,
PRIOR FILING DATE: 1999-02-01
PRIOR APPLICATION NUMBER: 60/126,593
PRIOR FILING DATE: 1999-03-26
PRIOR APPLICATION NUMBER: 60/134,093
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 60/134,092
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 60/165,124
PRIOR FILING DATE: 1999-11-12
PRIOR APPLICATION NUMBER: 60/165,086
PRIOR FILING DATE: 1999-11-12
NUMBER OF SEQ ID NOS: 292
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 259
LENGTH: 1289
TYPE: PRT
ORGANISM: *Mycobacterium tuberculosis*
US-09-712-363-259

Query Match 1.0%; Score 8; DB 10; Length 1289;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 SELAAEA 210
|||||
DB 756 SELAAEA 763

RESULT 48
US-10-128-714-8262
Sequence 8262, Application US/10128714
Publication No. US20030119013A1
GENERAL INFORMATION:
APPLICANT: Jiang, Bo
APPLICANT: Hu, Wengji
APPLICANT: Tishkoff, Daniel
APPLICANT: Zamudio, Carlos
APPLICANT: Eroshkin, Alexey M
APPLICANT: Lemieux, Sebastien M
TITLE OF INVENTION: Identification of Essential Genes in *Aspergillus fumigatus* and
TITLE OF INVENTION: Methods of Use
FILE REFERENCE: 10182-018-999
CURRENT APPLICATION NUMBER: US/10/128,714
CURRENT FILING DATE: 2002-04-23
PRIOR APPLICATION NUMBER: US 60/285,697
PRIOR FILING DATE: 2001-04-23
PRIOR APPLICATION NUMBER: US 60/287,066
PRIOR FILING DATE: 2001-04-27
PRIOR APPLICATION NUMBER: US 60/295,890
PRIOR FILING DATE: 2001-06-05
PRIOR APPLICATION NUMBER: US 60/303,899
PRIOR FILING DATE: 2001-07-09
PRIOR APPLICATION NUMBER: US 60/316,362
PRIOR FILING DATE: 2001-08-31

```

; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 8262
; LENGTH: 1353
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-8262
```

```

Query Match 1.0%; Score 8; DB 15; Length 1353;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 707 EEPVEETP 714
Db 1201 EEPVEETP 1208
```

```

RESULT 49
US-09-884-465A-23
; Sequence 23, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 23
; LENGTH: 41
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-23
```

```

Query Match 0.9%; Score 7; DB 11; Length 41;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 666 KYVEHP 672
Db 19 KYVEHP 25
```

```

RESULT 50
US-09-864-761-38022
; Sequence 38022, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wenheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemica-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-08-03
```

```

; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO: 38022
; LENGTH: 49
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC010734.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.93
US-09-864-761-38022
```

```

Query Match 0.9%; Score 7; DB 9; Length 49;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 492 LAPITHP 498
Db 34 LAPITHP 40
```

```

RESULT 51
US-10-106-698-5438
; Sequence 5438, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide.
; FILE REFERENCE: PA005P1
; CURRENT FILING DATE: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO: 5438
; LENGTH: 81
; TYPE: PRT
; ORGANISM: Homo sapiens
```


FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (72)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-5438

Query Match 0.9%; Score 7; DB 15; Length 81;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 787 PSSVSKE 793
|||||
DB 57 PSSVSKE 63

RESULT 52
US-09-764-891-3850
Sequence 3850, Application US/09764891
Publication No. US20030077808A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC006
CURRENT APPLICATION NUMBER: US/09/764,891
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 10231
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3850
LENGTH: 91
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (74)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-891-3850

Query Match 0.9%; Score 7; DB 11; Length 91;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 729 AQLKEAE 735
|||||
DB 21 AQLKEAE 27

RESULT 53
US-10-103-313-297
Sequence 297, Application US/10103313
Publication No. US20030082758A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PJ207C1
CURRENT APPLICATION NUMBER: US/10/103,313
CURRENT FILING DATE: 2002-03-12
NUMBER OF SEQ ID NOS: 653
Prior application removed - See File Wrapper or Palm
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 297
LENGTH: 97
TYPE: PRT
ORGANISM: Homo sapiens
US-10-103-313-297

Query Match 0.9%; Score 7; DB 15; Length 97;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 470 LIERLND 476
|||||
DB 68 LIERLND 74

RESULT 54
US-10-103-313-467
Sequence 467, Application US/10103313
Publication No. US20030082758A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PJ207C1
CURRENT APPLICATION NUMBER: US/10/103,313
CURRENT FILING DATE: 2002-03-12
NUMBER OF SEQ ID NOS: 653
Prior application removed - See File Wrapper or Palm
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 467
LENGTH: 97
TYPE: PRT
ORGANISM: Homo sapiens
US-10-103-313-467

Query Match 0.9%; Score 7; DB 15; Length 97;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 470 LIERLND 476
|||||
DB 68 LIERLND 74

RESULT 55
US-09-712-363-272
Sequence 272, Application US/09712363
Patent No. US20020164588A1
GENERAL INFORMATION:
APPLICANT: Eisenberg, David
APPLICANT: Rotstein, Sergio H.
APPLICANT: Marcotte, Edward M.
TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
FILE REFERENCE: 07419-032001
CURRENT APPLICATION NUMBER: US/09/712,363
CURRENT FILING DATE: 2000-11-13
Prior application number: PCT/US00/02246
Prior Filing Date: 2000-01-28
Prior Application Number: 60/179,531
Prior Filing Date: 2000-02-01
Prior Application Number: 60/117,844
Prior Filing Date: 1999-01-29
Prior Application Number: 60/118,206,
Prior Filing Date: 1999-02-01
Prior Application Number: 60/126,593
Prior Filing Date: 1999-03-26
Prior Application Number: 60/134,093
Prior Filing Date: 1999-05-14
Prior Application Number: 60/134,092
Prior Filing Date: 1998-05-14
Prior Application Number: 60/165,124
Prior Filing Date: 1999-11-12
Prior Application Number: 60/165,086
Prior Filing Date: 1999-11-12
NUMBER OF SEQ ID NOS: 292
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 272
LENGTH: 100
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
US-09-712-363-272

Query Match 0.9%; Score 7; DB 10; Length 100;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 605 GEXRIPL 611
Db 55 GEXRIPL 61

RESULT 56
US-10-252-945-57
; Sequence 57, Application US/10252945
; Publication No. US20030134904A1
; GENERAL INFORMATION:
; APPLICANT: Giordano, Tony
; APPLICANT: Sturgess, Michael A.
; APPLICANT: Rao, Samala, J.
; TITLE OF INVENTION: Inhibitors of RNASE P Proteins as
; FILE REFERENCE: 50093/026002
; CURRENT APPLICATION NUMBER: US/10/252,945
; CURRENT FILING DATE: 2002-09-23
; PRIOR APPLICATION NUMBER: US 60/323,853
; PRIOR FILING DATE: 2001-09-21
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 57
; LENGTH: 115
; TYPE: PRT
; ORGANISM: Staphylococcus epidermis
US-10-252-945-57

Query Match 0.9%; Score 7; DB 12; Length 115;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 460 RNSDFOA 466
Db 9 RNSDFOA 15

RESULT 57
US-10-156-761-7659
; Sequence 7659, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7659
; LENGTH: 139
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-7659

Query Match 0.9%; Score 7; DB 15; Length 139;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 733 EAEVILA 739
Db 130 EAEVILA 136

RESULT 58
US-09-820-843A-94
; Sequence 94, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEI
; FILE REFERENCE: 063915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 94
; LENGTH: 143
; TYPE: PRT
; ORGANISM: R. prowazekii
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: unknown
; NAME/KEY: misc feature
; OTHER INFORMATION: g1|3860651
US-09-820-843A-94

Query Match 0.9%; Score 7; DB 11; Length 143;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 133 NRQOEH 139
Db 106 NRQOEH 112

RESULT 59
US-09-817-414-2
; Sequence 2, Application US/09817414
; Patent No. US20020058308A1
; GENERAL INFORMATION:
; APPLICANT: Greenwood, Rebecca C.
; TITLE OF INVENTION: NOVEL grea
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESS: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/817,414
; FILING DATE: 26-Mar-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/229,804
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd O
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10046
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 158 amino acids
; TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-817-414-2

Query Match 0.9%; Score 7; DB 9; Length 158;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 IIEDTGD 182
Db 77 IIEDTGD 83

RESULT 60
US-10-043-487-355
; Sequence 355, Application US/10043487
; Publication No. US2003005220A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; APPLICANT: Pierre, LEGRAIN
; TITLE OF INVENTION: Protein-protein interactions between Shigella flexneri polypeptides
; FILE REFERENCE: B4778A
; CURRENT APPLICATION NUMBER: US/10/043,487
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/261,130
; PRIOR FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 561
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 355
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Shigella flexneri
US-10-043-487-355

Query Match 0.9%; Score 7; DB 15; Length 158;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 776 EKLAL 782
Db 3 EKLAL 9

RESULT 61
US-09-738-626-6669
; Sequence 6669, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASARO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patentin ver. 3.0
; SEQ ID NO 6669

LENGTH: 166
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-6669

Query Match 0.9%; Score 7; DB 10; Length 166;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 259 SQASQSN 265
Db 55 SQASQSN 61

RESULT 62
US-09-882-529-5
; Sequence 5, Application US/09882529
; Patent No. US20020132317A1
; GENERAL INFORMATION:
; APPLICANT: Peyman, John A
; APPLICANT: da Silva, Antonio
; TITLE OF INVENTION: NOVEL INTERFERON-INDUCED TETRASPAN PROTEIN AND NUCLEIC
; FILE REFERENCE: 15966-771
; CURRENT APPLICATION NUMBER: US/09/882,529
; PRIOR FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: 60/211,565
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 5
; LENGTH: 169
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-882-529-5

Query Match 0.9%; Score 7; DB 10; Length 169;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 759 NNLTLOI 765
Db 3 NNLTLOI 9

RESULT 63
US-09-882-529-6
; Sequence 6, Application US/09882529
; Patent No. US20020132317A1
; GENERAL INFORMATION:
; APPLICANT: Peyman, John A
; APPLICANT: da Silva, Antonio
; APPLICANT: Hockman, Paula
; TITLE OF INVENTION: NOVEL INTERFERON-INDUCED TETRASPAN PROTEIN AND NUCLEIC
; FILE REFERENCE: 15966-771
; CURRENT APPLICATION NUMBER: US/09/882,529
; PRIOR FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: 60/211,565
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 6
; LENGTH: 170
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-882-529-6

Query Match 0.9%; Score 7; DB 10; Length 170;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 773 AEAEXKL 779
|||
Db 91 AEAEXKL 97

RESULT 64

US-09-882-529-8
; Sequence 8, Application US/09882529
; Patent No. US2002013217A1
; GENERAL INFORMATION:
; APPLICANT: Peyman, John A
; APPLICANT: da Silva, Antonio
; APPLICANT: Hockman, Paula
; TITLE OF INVENTION: NOVEL INTERFERON-INDUCED TETRASPAN PROTEIN AND NUCLEIC
; FILE REFERENCE: 15966-771
; CURRENT APPLICATION NUMBER: US/09/882,529
; PRIOR FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: 60/211,565
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 170
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-882-529-8

Query Match 0.9%; Score 7; DB 15; Length 170;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 773 AEAEXKL 779
|||
Db 91 AEAEXKL 97

RESULT 65

US-10-156-761-9896
; Sequence 9896, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9896
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9896

Query Match 0.9%; Score 7; DB 15; Length 178;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 515 VRIAOA 521
|||
Db 133 VRIAOA 139

RESULT 66

US-10-267-311-35
; Sequence 35, Application US/10267311
; Publication No. US20030050469A1
; GENERAL INFORMATION:
; APPLICANT: Siegel, Marvin
; APPLICANT: Chu, N. Randall
; APPLICANT: Mizzen, Lee A.
; TITLE OF INVENTION: INDUCTION OF A TH1-LIKE RESPONSE IN VITRO
; FILE REFERENCE: 12071/002001
; CURRENT APPLICATION NUMBER: US/10/267,311
; CURRENT FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: US/09/613,303
; PRIOR FILING DATE: 2000-07-10
; PRIOR APPLICATION NUMBER: US 60/143,757
; PRIOR FILING DATE: 1999-07-08
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 198
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fusion sequence
US-10-267-311-35

Query Match 0.9%; Score 7; DB 15; Length 198;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 605 GEKRIPL 611
|||
Db 55 GEKRIPL 61

RESULT 67

US-09-967-736-3
; Sequence 3, Application US/09967736
; Patent No. US20020103340A1
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: RAB PROTEINS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESS: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/967,736
; FILING DATE: 28-Sep-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/154,602
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0367 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:

LENGTH: 201 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: LVRVTUT04
CLONE: 2514506
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-967-736-3

Query Match 0.9%; Score 7; DB 10; Length 201;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 365 PNLKIDS 371
Db 184 PNLKIDS 190

RESULT 68

US-09-967-736-8
Sequence 8, Application US/09967736
Patent No. US20020103340A1
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
Lal, Preeti
Corley, Neil C.
Shah, Purvi

TITLE OF INVENTION: RAB PROTEINS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible

OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/967,736
FILING DATE: 28-Sep-2001

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/154,602
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0367 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 201 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 57006

SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-967-736-8

Query Match 0.9%; Score 7; DB 10; Length 201;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 365 PNLKIDS 371
Db 184 PNLKIDS 190

Db 184 PNLKIDS 190

RESULT 69

US-10-156-761-11020
Sequence 11020, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OKURA, SATOSHI
APPLICANT: IKEDA, HARUO

APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761

PRIOR FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30

PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109

SEQ ID NO 11020
LENGTH: 209
TYPE: PRT

ORGANISM: Streptomyces avermitilis
US-10-156-761-11020

Query Match 0.9%; Score 7; DB 15; Length 209;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 488 LIAFLAP 494
Db 100 LIAFLAP 106

RESULT 70

US-09-925-301-1048
Sequence 1048, Application US/09925301
Patent No. US20020052308A1
GENERAL INFORMATION:

APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA106
CURRENT APPLICATION NUMBER: US/09/925,301

PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05882
PRIOR FILING DATE: 2000-03-08

PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1694

SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1048
LENGTH: 217
TYPE: PRT

ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (122)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (186)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (200)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-1048

Query Match 0.9%; Score 7; DB 9; Length 217;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 127 RTKEIN 133
Db 72 RTKEIN 78

RESULT 71

US-10-102-806-466
; Sequence 466, Application US/10102806
; Publication No. US20030054421A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA103PIC1
; CURRENT APPLICATION NUMBER: US/10/102,806
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/925,298
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05881
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 846
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 466
; LENGTH: 224
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-102-806-466

Query Match 0.9%; Score 7; DB 15; Length 224;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 365 PNLKIDS 371
Db 207 PNLKIDS 213

RESULT 72

US-09-604-287A-487
; Sequence 487, Application US/09604287A
; Patent No. US20020064872A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Dillon, Devin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C7
; CURRENT APPLICATION NUMBER: US/09/604,287A
; CURRENT FILING DATE: 2000-06-22
; NUMBER OF SEQ ID NOS: 489
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 487
; LENGTH: 235
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-604-287A-487

Query Match 0.9%; Score 7; DB 9; Length 235;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 127 RTKEIN 133
Db 47 RTKEIN 53

RESULT 73
US-09-907-263-4

; Sequence 4, Application US/09907263
; Patent No. US20020119924A1
; GENERAL INFORMATION:
; APPLICANT: Bendele, Alison M.
; APPLICANT: Semello, Regina M.
; APPLICANT: Edwards, Carl K.
; TITLE OF INVENTION: COMBINATION THERAPY USING A TNF BINDING
; PROTEIN FOR TREATING TNF-MEDIATED DISEASES
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 DeHavilland Drive
; CITY: Thousand Oaks
; STATE: CA
; COUNTRY: US
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/907,263
; FILING DATE: 17-Jul-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/326,394
; FILING DATE: 1999-06-04
; APPLICATION NUMBER: US 60/036,355
; FILING DATE: 23-JAN-1997
; APPLICATION NUMBER: US 60/039,315
; FILING DATE: 07-FEB-1997
; APPLICATION NUMBER: US 60/052,023
; FILING DATE: 09-JUL-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Zindrick, Thomas K.
; REGISTRATION NUMBER: 32,185
; REFERENCE/DOCKET NUMBER: A-430D
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 235 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-907-263-4

Query Match 0.9%; Score 7; DB 10; Length 235;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 352 QPTPEPS 358
Db 206 QPTPEPS 212

RESULT 74

US-09-882-735-16
; Sequence 16, Application US/09882735
; Publication No. US20030054439A1
; GENERAL INFORMATION:
; APPLICANT: Fisher F., Eric
; APPLICANT: Edwards K., Carl
; APPLICANT: Kieft L., Gary
; TITLE OF INVENTION: Truncated Soluble Tumor Necrosis Factor Type-I and
; TITLE OF INVENTION: Type-II Receptors
; FILE REFERENCE: 02-006
; CURRENT APPLICATION NUMBER: US/09/882,735
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 09/214,613
; PRIOR FILING DATE: 1999-01-08

```

; PRIOR APPLICATION NUMBER: PCT/US97/12244
; PRIOR FILING DATE: 1997-07-09
; PRIOR APPLICATION NUMBER: 60/039,792
; PRIOR FILING DATE: 1997-03-04
; PRIOR APPLICATION NUMBER: 60/039,314
; PRIOR FILING DATE: 1997-02-07
; PRIOR APPLICATION NUMBER: 60/037,737
; PRIOR FILING DATE: 1997-01-23
; PRIOR APPLICATION NUMBER: 60/032,354
; PRIOR FILING DATE: 1996-12-06
; PRIOR APPLICATION NUMBER: 60/021,443
; PRIOR FILING DATE: 1996-07-09
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 235
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-882-735-16

```

```

Query Match          0.9%; Score 7; DB 11; Length 235;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      352 QPTPEPS 358
Db      206 QPTPEPS 212

```

```

RESULT 75
US-10-124-805-487
; Sequence 487, Application US/10124805
; Publication No. US20030166022A1
; GENERAL INFORMATION:
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: Persing, David H.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C12
; CURRENT APPLICATION NUMBER: US/10/124,805
; CURRENT FILING DATE: 2002-04-15
; NUMBER OF SEQ ID NOS: 627
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 487
; LENGTH: 235
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-124-805-487

```

```

Query Match          0.9%; Score 7; DB 12; Length 235;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      127 RTKEIN 133
Db      47 RTKEIN 53

```

```

RESULT 76
US-10-436-826-75
; Sequence 75, Application US/10436826
; Publication No. US20030187224A1
; GENERAL INFORMATION:
; APPLICANT: Boyle, William J.
; APPLICANT: Wooden, Scott
; FILE REFERENCE: 06843.0034-01000
; CURRENT APPLICATION NUMBER: US/10/436,826
; CURRENT FILING DATE: 2003-05-12
; PRIOR APPLICATION NUMBER: 08/850,188
; PRIOR FILING DATE: 1997-05-01
; NUMBER OF SEQ ID NOS: 88

```

```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 75
; LENGTH: 235
; TYPE: PRT
; ORGANISM: Human
US-10-436-826-75

```

```

Query Match          0.9%; Score 7; DB 12; Length 235;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      352 QPTPEPS 358
Db      206 QPTPEPS 212

```

```

RESULT 77
US-10-007-805-487
; Sequence 487, Application US/10007805
; Publication No. US20020150581A1
; GENERAL INFORMATION:
; APPLICANT: Uiang, Yugu
; APPLICANT: Dillon, Devin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jianshun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepner, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Durham, Margareta
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C10
; CURRENT APPLICATION NUMBER: US/10/007,805
; CURRENT FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 593
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 487
; LENGTH: 235
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-007-805-487

```

```

Query Match          0.9%; Score 7; DB 14; Length 235;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      127 RTKEIN 133
Db      47 RTKEIN 53

```

```

RESULT 78
US-10-076-622-487
; Sequence 487, Application US/10076622
; Publication No. US20030023036A1
; GENERAL INFORMATION:
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: Persing, David H.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C11
; CURRENT APPLICATION NUMBER: US/10/076,622
; CURRENT FILING DATE: 2002-02-13
; NUMBER OF SEQ ID NOS: 627
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 487
; LENGTH: 235
; TYPE: PRT
; ORGANISM: Homo sapiens

```

US-10-076-622-487

Query Match 0.9%; Score 7; DB 15; Length 235;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 RTKEIN 133
|||||
47 RTKEIN 53

RESULT 79

US-10-243-230-2
; Sequence 2, Application US/10243230
; Publication No. US20030023049A1

GENERAL INFORMATION:

APPLICANT: Pettit, Dean

TITLE OF INVENTION: Site Specific Protein Modification

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESSES:

STREET: 51 University

CITY: Seattle

STATE: WA

COUNTRY: US

ZIP: 98101

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/243,230

FILING DATE: 12-Sep-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/102,530

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Henry, Janis C

REGISTRATION NUMBER: 34,347

REFERENCE/DOCKET NUMBER: 2637

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206)470-4189

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 235 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-10-243-230-2

Query Match 0.9%; Score 7; DB 15; Length 235;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 QTPPEPS 358
|||||
206 QTPPEPS 212

RESULT 80

US-10-243-230-4
; Sequence 4, Application US/10243230
; Publication No. US20030023049A1

GENERAL INFORMATION:

APPLICANT: Pettit, Dean

TITLE OF INVENTION: Site Specific Protein Modification

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESSES:

STREET: 51 University

CITY: Seattle

STATE: WA

COUNTRY: US

ZIP: 98101

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/243,230

FILING DATE: 12-Sep-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/102,530

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Henry, Janis C

REGISTRATION NUMBER: 34,347

REFERENCE/DOCKET NUMBER: 2637

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206)470-4189

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 235 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Query Match 0.9%; Score 7; DB 15; Length 235;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 QTPPEPS 358
|||||
206 QTPPEPS 212

RESULT 81

US-10-243-230-6
; Sequence 6, Application US/10243230
; Publication No. US20030023049A1

GENERAL INFORMATION:

APPLICANT: Pettit, Dean

TITLE OF INVENTION: Site Specific Protein Modification

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESSES:

STREET: 51 University

CITY: Seattle

STATE: WA

COUNTRY: US

ZIP: 98101

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/243,230

FILING DATE: 12-Sep-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/102,530

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Henry, Janis C

REGISTRATION NUMBER: 34,347

REFERENCE/DOCKET NUMBER: 2637

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206)470-4189

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 235 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-243-230-6

Query Match
Best Local Similarity 100.0%; Score 7; DB 15; Length 235;
Pred. No. 3e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

Qy 352 QPTPEPS 358
Db 206 QPTPEPS 212

RESULT 82
US-10-243-230-8
Sequence 8, Application US/10243230
Publication No. US20030023049A1

GENERAL INFORMATION:
APPLICANT: Pettit, Dean
TITLE OF INVENTION: Site Specific Protein Modification
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Janis C. Henry
STREET: 51 University
CITY: Seattle
STATE: WA
COUNTRY: US
ZIP: 98101

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/243,230
FILING DATE: 12-Sep-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/102,530
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Henry, Janis C
REGISTRATION NUMBER: 34,347
REFERENCE/DOCKET NUMBER: 2637
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)470-4189

INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 235 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-10-243-230-8

Query Match
Best Local Similarity 100.0%; Score 7; DB 15; Length 235;
Pred. No. 3e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

Qy 352 QPTPEPS 358
Db 206 QPTPEPS 212

RESULT 83
US-10-156-761-12784
Sequence 12784, Application US/10156761
Publication No. US20030119018A1

GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 12784
LENGTH: 247
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-12784

Query Match
Best Local Similarity 100.0%; Score 7; DB 15; Length 247;
Pred. No. 3.1e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

Qy 487 DLLAFIA 493
Db 164 DLLAFIA 170

RESULT 84
US-10-032-585-7603
Sequence 7603, Application US/10032585
Publication No. US20030180953A1
GENERAL INFORMATION:
APPLICANT: Terry, Roemer D.
APPLICANT: Bo, Jiang
APPLICANT: Charles, Boone
APPLICANT: Howard, Bussey
TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
FILE REFERENCE: 10182-005-999
CURRENT APPLICATION NUMBER: US/10/032,585
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 8000
SOFTWARE: PatentIn version 3.1
SEQ ID NO 7603
LENGTH: 248
TYPE: PRT
ORGANISM: Candida albicans
US-10-032-585-7603

Query Match
Best Local Similarity 100.0%; Score 7; DB 12; Length 248;
Pred. No. 3.2e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

Qy 373 SSLVSQL 379
Db 27 SSLVSQL 33

RESULT 85
US-10-313-852-10
Sequence 10, Application US/10313852
Publication No. US20030103942A1
GENERAL INFORMATION:
APPLICANT: Burslein, Haim
APPLICANT: Stepan, Anthony M.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR LOWERING
THE LEVEL OF TUMOR NECROSIS FACTOR (TNF) IN TNF-ASSOCIATED
DISORDERS
FILE REFERENCE: 226272004420

```

; CURRENT APPLICATION NUMBER: US/10/313,852
; CURRENT FILING DATE: 2002-12-06
; PRIOR APPLICATION NUMBER: US/09/579,845
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/150,688
; PRIOR FILING DATE: 1999-05-28
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 257
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-313-852-10

```

```

Query Match      0.9%; Score 7; DB 15; Length 257;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      352 QTPPEPS 358
        |||||
Db      228 QTPPEPS 234

```

```

RESULT 86
US-10-314-033-10
; Sequence 10, Application US/10314033
; Publication No. US20030113295A1
; GENERAL INFORMATION:
; APPLICANT: Burstein, Haim
; APPLICANT: Stepan, Anthony M.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR LOWERING
; TITLE OF INVENTION: THE LEVEL OF TUMOR NECROSIS FACTOR (TNF) IN TNF-ASSOCIATED
; TITLE OF INVENTION: DISORDERS
; FILE REFERENCE: 226272004420
; CURRENT APPLICATION NUMBER: US/10/314,033
; CURRENT FILING DATE: 2002-12-06
; PRIOR APPLICATION NUMBER: US/09/579,845
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/150,688
; PRIOR FILING DATE: 1999-05-28
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 257
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-314-033-10

```

```

Query Match      0.9%; Score 7; DB 15; Length 257;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      352 QTPPEPS 358
        |||||
Db      228 QTPPEPS 234

```

```

RESULT 87
US-10-156-761-11174
; Sequence 1174, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29

```

```

; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 11174
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-11174

```

```

Query Match      0.9%; Score 7; DB 15; Length 264;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      200 LSASELA 206
        |||||
Db      142 LSASELA 148

```

```

RESULT 88
US-09-884-465A-304
; Sequence 304, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Joesee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 304
; LENGTH: 266
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-304

```

```

Query Match      0.9%; Score 7; DB 11; Length 266;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      598 AIYNRVK 604
        |||||
Db      74 AIYNRVK 80

```

```

RESULT 89
US-09-884-465A-303
; Sequence 303, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Joesee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20

```

```

; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 303
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-303
```

```

Query Match          0.9%; Score 7; DB 11; Length 268;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      598 AIYNRVK 604
      |||||
Db       74 AIYNRVK 80
```

```

RESULT 90
US-09-884-465A-375
; Sequence 375, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 375
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-375
```

```

Query Match          0.9%; Score 7; DB 11; Length 268;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      598 AIYNRVK 604
      |||||
Db       74 AIYNRVK 80
```

```

RESULT 91
US-09-884-465A-294
; Sequence 294, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
```

```

; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 294
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-294
```

```

Query Match          0.9%; Score 7; DB 11; Length 272;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      598 AIYNRVK 604
      |||||
Db       74 AIYNRVK 80
```

```

RESULT 92
US-09-884-465A-295
; Sequence 295, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 295
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-295
```

```

Query Match          0.9%; Score 7; DB 11; Length 272;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      598 AIYNRVK 604
      |||||
Db       74 AIYNRVK 80
```

```

RESULT 93
US-09-884-465A-296
; Sequence 296, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
```

```
; SEQ ID NO 296
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-296
```

```
Query Match
Best Local Similarity 100.0%; Score 7; DB 11; Length 272;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 598 AIYNRVK 604
|||
Db 74 AIYNRVK 80
```

```
RESULT 94
US-09-884-465A-297
; Sequence 297, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 297
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-297
```

```
Query Match
Best Local Similarity 100.0%; Score 7; DB 11; Length 272;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 598 AIYNRVK 604
|||
Db 74 AIYNRVK 80
```

```
RESULT 95
US-09-884-465A-298
; Sequence 298, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 298
```

```
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-298
```

```
Query Match
Best Local Similarity 100.0%; Score 7; DB 11; Length 272;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 598 AIYNRVK 604
|||
Db 74 AIYNRVK 80
```

```
RESULT 96
US-09-884-465A-299
; Sequence 299, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 299
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-299
```

```
Query Match
Best Local Similarity 100.0%; Score 7; DB 11; Length 272;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 598 AIYNRVK 604
|||
Db 74 AIYNRVK 80
```

```
RESULT 97
US-09-884-465A-301
; Sequence 301, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 301
; LENGTH: 272
```

TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism
US-09-884-465A-301

Query Match 0.9%; Score 7; DB 11; Length 272;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 598 AAYNRK 604
DB 74 AAYNRK 80

RESULT 98
US-10-017-161-910
Sequence 910, Application US/10017161
Publication No. US20030143668A1
GENERAL INFORMATION:
APPLICANT: SUMA, MAKIKO
APPLICANT: ASAI, KIYOSHI
APPLICANT: AKIYAMA, YUTAKA
APPLICANT: ABURATANI, HIROYUKI
TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
FILE REFERENCE: 084335/0152
CURRENT APPLICATION NUMBER: US/10/017,161
CURRENT FILING DATE: 2002-12-18
PRIOR APPLICATION NUMBER: JP 2001/246789
PRIOR FILING DATE: 2001-06-18
NUMBER OF SEQ ID NOS: 2430
SOFTWARE: Patentin Ver. 2.1
SEO ID NO 910
LENGTH: 285
TYPE: PRT
ORGANISM: Homo sapiens
US-10-017-161-910

Query Match 0.9%; Score 7; DB 12; Length 285;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 733 EAEVLLA 739
DB 87 EAEVLLA 93

RESULT 99
US-10-032-585-7197
Sequence 7197, Application US/10032585
Publication No. US20030180953A1
GENERAL INFORMATION:
APPLICANT: Terry, Roemer D.
APPLICANT: Bo, Jlang
APPLICANT: Charles, Boone
APPLICANT: Howard, Bussey
TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
FILE REFERENCE: 10182-005-999
CURRENT APPLICATION NUMBER: US/10/032,585
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 8000
SOFTWARE: Patentin version 3.1
SEO ID NO 7197
LENGTH: 293
TYPE: PRT
ORGANISM: Candida albicans
US-10-032-585-7197

Query Match 0.9%; Score 7; DB 12; Length 293;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 773 AEAEXLL 779

DB 9 AEAEXLL 15

RESULT 100
US-09-886-055-113
Sequence 113, Application US/09886055
Patent No. US20020132273A1
GENERAL INFORMATION:
APPLICANT: SRYVER, LUBERT
APPLICANT: ZOZULIA, SERGEY
TITLE OF INVENTION: RECEPTOR FINGERPRINTING, SENSORY PERCEPTION, AND
TITLE OF INVENTION: BIOSENSORS OF CHEMICAL SENSANTS
FILE REFERENCE: 078003-0277150
CURRENT APPLICATION NUMBER: US/09/886,055
CURRENT FILING DATE: 2001-06-22
PRIOR APPLICATION NUMBER: 60/213,812
PRIOR FILING DATE: 2000-06-22
NUMBER OF SEQ ID NOS: 522
SOFTWARE: Patentin Ver. 2.1
SEO ID NO 113
LENGTH: 311
TYPE: PRT
ORGANISM: Homo sapiens
US-09-886-055-113

Query Match 0.9%; Score 7; DB 10; Length 311;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 733 EAEVLLA 739
DB 113 EAEVLLA 119

Search completed: November 14, 2003, 11:38:27
Job time : 79 secs

THIS PAGE BLANK (USPTO)